Stockholms universitet

Research Report

Department of Statistics

No. 2012:1

Random Stub Matching Models of Multigraphs

Termeh Shafie

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

Random Stub Matching Models of Multigraphs

Termeh Shafie

Abstract

This article studies the local and global structure of multigraphs under random stub matching with fixed degrees (RSM). The local structure is analyzed by marginal distributions of edge multiplicities, and the global structure is analyzed by the simultaneous distribution of edge multiplicities. The simultaneous distribution is shown to depend on a single complexity statistic. The distributions under RSM and IEA are used for calculations of moments and entropies, and for comparisons by information divergence. The modified distributions are obtained by ignoring the dependencies between edges and assuming independent edge assignments to sites (IEA), and by ignoring the dependencies between stubs and assuming independent stub assignments to vertices (ISA). The main results in this article include a new formula for the probability of an arbitrary number of loops at a vertex, and a more intricate expression for the probability of an arbitrary number of edges at any site. Further, simplicity and complexity of multigraphs under RSM are investigated and a new method of approximating the probability that an RSM multigraph is simple is proposed and shown to perform well for multigraphs with small numbers of vertices and edges.

Keywords: multigraph, edge multiplicity, entropy, information divergence, simplicity and complexity.

1 Introduction

It is well known that different degree sequences are compatible with different numbers of graphs. Several methods have been developed for generating random graphs with fixed or modeled degrees, degree distributions or expected degrees. Such methods can be found in Blitzstein and Diaconis (2011), Bayati, Kim and Saberi (2010), Britton, Deijfen and Martin-Löf (2006), Chung and Lu (2002), and Bender and Canfield (1978). Random stub matching, also referred to as the configuration model or the pairing model (e.g. Janson 2009, Bollobàs 1980), generates random multigraphs by randomly coupling pairs of stubs to form edges.

Department of Statistics, Stockholm University, S-106 91 Stockholm, termeh.shafie@stat.su.se

This article focuses on both the local and global structure of multigraphs under random stub matching with fixed degrees (RSM). The local structure is analyzed by marginal distributions of edge multiplicities, and the global structure is analyzed by the simultaneous distribution of edge multiplicities. The distributions under RSM as well as some modified distributions with modeled degrees are used for calculations of moments and entropies, and for comparisons by information divergences. These modified distributions are obtained by ignoring the dependencies between edges when they are assigned to sites (IEA) and by ignoring the dependencies between stubs when they are assigned to vertices (ISA).

In the next section, some basic concepts such as stubs, edges, sites, multigraphs, multiplicities and complexities are presented. The uniform random stub matching procedure given fixed degrees is described in Section 3 where the distribution of multigraphs is determined and shown to depend on a single statistic which is a special summary measure of complexity.

The moments of the edge multiplicity distributions under RSM are derived in Section 4. The moments of the number of loops at a fixed vertex are determined as functions of the number of edges, denoted m, and the degree at that vertex. It is shown that the variance of the number of loops under RSM is less than the variance under IEA, except for the degenerate cases of degree value 1 or $2m$. The moments of the number of edges between two distinct vertices are determined as functions of the total number of edges m and the degrees at the two vertices. It is shown that the variance of the number of such edges under RSM is generally less than the variance under IEA, except for degrees that lie symmetrically around the total number of edges and are given by $m \pm k$ for any non-negative integer k less than a specified limit.

In Section 5, the distributions of edge multiplicities at local sites are investigated. The probability of no loops at a vertex has been given in the literature (Janson 2009) but, as far as we know, not generalized to arbitrary number of loops at a vertex. This formula is derived using a special technique which also allows us to find the probability of edge frequency between pairs of distinct vertices. This technique gives the trivariate distribution of the numbers of loops and non-loops within and between two distinct vertices, and from its marginals we obtain distributions of local edge multiplicities at any site. Although somewhat combinatorically tedious, we also determine the range space of the trivariate distribution.

Throughout Section 5, information divergence and entropies are used to compare the edge multiplicity distributions under RSM and IEA. Numerical examples using the divergence indicate that the distribution of loop multiplicity under RSM is closely related to that of the IEA distribution at vertices with low degrees. The divergence increases monotonically from zero to a maximal value and decreases very steeply back to zero. The results also indicate that the discrepancy between the edge multiplicity distributions under RSM and IEA is due to their different range spaces. Further, an illustration is given of how the divergence between the probability distributions of local edge frequency under RSM and

IEA varies for different degrees. The results also show how the resemblance between the distributions increases with increasing m.

The flatness of the local edge multiplicity distributions under RSM and IEA are compared using entropies. Specifically, the entropy of the loop multiplicity distribution under RSM is shown to be more symmetrical than that of IEA, and it has its maximum around the stub proportion value 0.5. The corresponding loop multiplicity distribution under IEA is skew to the right and has its maximum for a stub proportion value of about 0.7. Good approximations to the entropies of loop multiplicities under both RSM and IEA are found.

Special attention is payed to the edge multiplicity distribution for the case with two degrees that lie symmetrically around m. For this case the entropies are much higher for IEA than for RSM. For both RSM and IEA, we give approximations to the entropies. These approximations are very good for the IEA distributions but not for the RSM distributions.

In Section 6 the global structure is analyzed by the distribution of multigraphs under RSM and IEA. Under RSM, this distribution was earlier shown to depend on a single complexity statistic, and in order to find the entropy of this distribution, results about edge multiplicities from Section 5 are used. The approximate entropies of the RSM and IEA distribution of multigraphs are given using covariance matrices. For both RSM and IEA, the exact and approximate entropies are close to the upper bounds of the exact entropies. Using the information divergence, a large deviation between the multigraph distributions under RSM and IEA are found and the results indicate flat distributions over very different ranges. In particular for regular multigraphs, both the RSM and IEA distribution cluster at the high probability sites when more edges are added and are therefore less flat for large values of m.

In the final section, the simplicity and complexity of multigraphs under RSM are studied. Two asymptotic results for the probability that an RSM multigraph is simple are numerically investigated, and it is shown that these probability approximations do not perform well for multigraphs with small numbers of vertices and edges. Under certain conditions, an alternative way of approximating the probability that an RSM multigraph is simple is proposed. Numerical examples show that this approximation is good for small multigraphs. Some other variables that identify simplicity and complexity are also considered, and the moments of these variables are derived. It is shown that the moments of some of these variables are much easier handled under IEA and a convenient way of obtaining the IEA distribution is introduced. This is done by assuming that the stubs are randomly generated and independently assigned to sites (ISA) and can be viewed as a Bayesian model for the stub frequencies under RSM. Using this method, approximations to the entropy of the distribution of multigraphs under RSM are derived. If the degree distributions are uniformly or close to uniformly distributed, the approximations are good even for small multigraphs, and for skew distributions they are good for multigraphs with many edges. An asymptotic equipartition property is shown to give alternative approximations that work reasonably well except for multigraphs with skew degree sequences and few vertices.

2 Basic Concepts and Notation

A finite undirected graph g with n labeled vertices and m labeled edges associates with each edge an ordered or unordered vertex pair. Let $V = \{1, \ldots, n\}$ and $E = \{1, \ldots, m\}$ be the sets of vertices and edges labeled by integers, and denote by R the set of available sites for the edges. The site space for directed edges is V^2 and the site space for undirected edges is $R = \{(i, j) \in V^2 : i \leq j\}$. We consider (i, j) with $i \leq j$ as a canonical representation for the unordered vertex pair. Let $r = \binom{n+1}{2}$ be the number of sites in R.

The degree of vertex i, the number of edges incident to it, is denoted d_i and $\sum_i^n d_i = 2m$. The degree sequence $\mathbf{d} = (d_1, \ldots, d_n)$ defines another sequence of $2m$ vertices or edge-stubs corresponding to m edges without specifying the pairings of stubs to edges:

$$\mathbf{s} = (\underbrace{1 \ldots 1}_{d_1} \quad \underbrace{2 \ldots 2}_{d_2} \quad \cdots \quad \underbrace{n \ldots n}_{d_n}) \ .$$

Thus there is a bijection $\mathbf{d} \leftrightarrow \mathbf{s}$ and we use the shorthand notation

$$\mathbf{s} = (s_1, \ldots, s_{2m}) = (1^{d_1} 2^{d_2} \cdots n^{d_n}) \in V^{2m} \ .$$

Let $X(\mathbf{d})$ be the set of sequences $\mathbf{x}$ that are permutations of the stub sequence

$$X(\mathbf{d}) = \{\mathbf{x} = (x_1, \ldots, x_{2m}) \in V^{2m} : \mathbf{x} \sim \mathbf{s}\} \ ,$$

where $\sim$ means "is a permutation of". The number of permutations of a stub sequence $\mathbf{s}$ obtained from the degree sequence $\mathbf{d}$ is given by

$$|X(\mathbf{d})| = \binom{2m}{\mathbf{d}} = \frac{(2m)!}{\mathbf{d}!} = \frac{(2m)!}{\prod_{i=1}^n d_i!} \ ,$$

and is also denoted $\#(\mathbf{x}|\mathbf{s})$ or $\#(\mathbf{x}|\mathbf{d})$.

Edges at the same site are called multiple edges, and the number of multiple edges at site (i, j) is its multiplicity denoted m_{ij}. The multiplicity at site $(i, j) \in V^2$ in $\mathbf{x}$ is

$$m_{ij}(\mathbf{x}) = \sum_{k=1}^{m} I\left((x_{2k-1}, x_{2k}) = (i, j)\right) \ .$$

It follows that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij}(\mathbf{x}) = m_{\cdot\cdot}(\mathbf{x}) = m$$

and

$$\sum_{j=1}^{n} (m_{ij}(\mathbf{x}) + m_{ji}(\mathbf{x})) = m_{i\cdot}(\mathbf{x}) + m_{\cdot i}(\mathbf{x}) = d_i \quad \text{for} \quad i = 1, \ldots, n \ .$$

The number of loops and the number of non-loops are denoted

$$m_1(\mathbf{x}) = \sum_{i=1}^{n} m_{ii}(\mathbf{x}) \qquad \text{and} \qquad m_2(\mathbf{x}) = \sum\sum_{i \neq j} m_{ij}(\mathbf{x}) \ .$$

When the edge multiplicities in $\mathbf{x}$ are arranged as a matrix we obtain the edge multiplicity matrix

$$\mathbf{m}(\mathbf{x}) = \big(m_{ij}(\mathbf{x}) : (i,j) \in V^2\big)$$

with loop counts $m_{ii}(\mathbf{x})$ in the main diagonal. If a matrix is created with these loop-counts as the elements in the main diagonal and zeros outside, we obtain the loop frequency matrix $\mathbf{m_1}(\mathbf{x})$. The non-loop frequency matrix denoted $\mathbf{m_2}(\mathbf{x})$ is then given as $\mathbf{m_2}(\mathbf{x}) = \mathbf{m}(\mathbf{x}) - \mathbf{m_1}(\mathbf{x})$.

The representation of the edge sequence is modified in two different ways. The sequence $\mathbf{y} = (y_1, \ldots, y_{2m})$ is obtained from $\mathbf{x}$ by vertex shifts according to $(y_{2k-1}, y_{2k}) = (\min(x_{2k-1}, x_{2k}), \max(x_{2k-1}, x_{2k}))$ for $k = 1, \ldots, m$. In other words, the injective map $\mathbf{x} \to \mathbf{y}$ gives an ordered sequence of m edges from R and $\mathbf{y}$ is the edge sequence of an undirected graph generated from $\mathbf{x}$. The number of $\mathbf{x}$ that yield the same $\mathbf{y}$ is then given as $\#(\mathbf{x}|\mathbf{y}) = 2^{m_2(\mathbf{y})}$. The sequence $\mathbf{z} = (z_1, \ldots, z_{2m})$ is obtained from $\mathbf{y}$ by ordering its edges non-decreasingly, i.e. the injective map $\mathbf{y} \to \mathbf{z}$ gives an edge sequence canonically ordered according to

$$(1,1) < (1,2) < \cdots < (1,n) < (2,2) < (2,3) < \cdots < (n,n)$$

so that

$$(z_1, z_2) \leq (z_3, z_4) \leq \cdots \leq (z_{2m-1}, z_{2m}) \ .$$

The edge sequence $\mathbf{z}$ represents the vertex labeled graph given by $\mathbf{y}$ without the edge labels. The set of sequences $\mathbf{z}$ generated by $\mathbf{x} \in X(\mathbf{d})$ is denoted $Z(\mathbf{d})$. The number of $\mathbf{x}$ that yield the same vertex labeled graph $\mathbf{z}$ is $\#(\mathbf{x}|\mathbf{z}) = \sum_{\mathbf{y}|\mathbf{z}} \#(\mathbf{x}|\mathbf{y}) = \sum_{\mathbf{y}|\mathbf{z}} 2^{m_2(\mathbf{y})}$. Now $m_2(\mathbf{y}) = m_2(\mathbf{z})$ and

$$\#(\mathbf{y}|\mathbf{z}) = \binom{m}{\mathbf{m}(\mathbf{z})} = \frac{m!}{\prod_{i \leq j} \ m_{ij}(\mathbf{z})!} \ ,$$

so that

$$\#(\mathbf{x}|\mathbf{z}) = 2^{m_2(\mathbf{z})} \binom{m}{\mathbf{m}(\mathbf{z})} \ .$$

Here the edge multiplicity at site $(i,j) \in V^2$ in $\mathbf{y}$ and $\mathbf{z}$ is equal to the common value

$$m_{ij}(\mathbf{y}) = m_{ij}(\mathbf{z}) = \begin{cases} m_{ij}(\mathbf{x}) & \text{for } i = j \\ m_{ij}(\mathbf{x}) + m_{ji}(\mathbf{x}) & \text{for } i < j \\ 0 & \text{for } i > j \ . \end{cases}$$

The common edge multiplicity matrices of $\mathbf{y}$ and $\mathbf{z}$ are triangular with zeros below the main diagonal. Moreover, the loop frequency matrices of $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are all equal. The numbers of loops and non-loops are therefore equal to

$$m_1(\mathbf{z}) = \sum_{i=1}^{n} m_{ii}(\mathbf{z}) = \sum_{i=1}^{n} m_{ii}(\mathbf{x})$$

and

$$m_2(\mathbf{z}) = \sum_{i<j} \sum m_{ij}(\mathbf{z}) = \sum \sum_{i \neq j} m_{ij}(\mathbf{x}) \ .$$

The sum of the multiplicity matrix and its transpose is the same symmetric matrix for $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$, i.e.

$$\mathbf{m}(\mathbf{x}) + \mathbf{m}'(\mathbf{x}) = \mathbf{m}(\mathbf{y}) + \mathbf{m}'(\mathbf{y}) = \mathbf{m}(\mathbf{z}) + \mathbf{m}'(\mathbf{z}) \ .$$

The row and column sums in this matrix are given by the degrees, and the loop counts are doubled in the main diagonal.

Figure 1 shows a schematic view of bijections and other functional relationships between the various concepts introduced here. The functional relationships comprise three different edge sequences and their edge multiplicity matrices, the stub sequence, and the degree sequence.

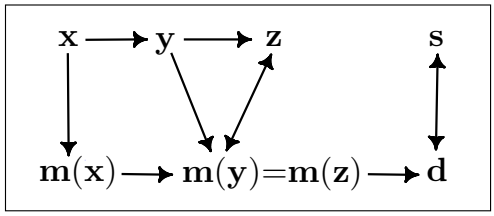


Figure 1: Functional relationships between the edge sequences, multiplicity matrices, stub sequence and degree sequence.

3 Uniform Stub Matching with Fixed Degrees

We focus on uniform distributions for different families of graphs which we refer to as random graphs. Assume that $\boldsymbol{\xi}$ is a random permutation of the stub sequence $\mathbf{s}$ defined by the degree sequence $\mathbf{d}$, i.e. $\boldsymbol{\xi}$ is uniform on $X(\mathbf{d}) = \{\mathbf{x} : \mathbf{x} \sim \mathbf{s}\} \subseteq V^{2m}$ with probabilities

$$P(\boldsymbol{\xi} = \mathbf{x}) = \frac{1}{\binom{2m}{\mathbf{d}}} \qquad \text{for} \quad \mathbf{x} \in X(\mathbf{d}) \ .$$

Let $\boldsymbol{\eta}$ be the edge sequence of the undirected graph obtained by shifts in $\boldsymbol{\xi}$. Further let $\boldsymbol{\zeta}$ be the canonical edge sequence of the undirected graph generated by $\boldsymbol{\xi}$ with probabilities

$$P(\boldsymbol{\zeta} = \mathbf{z}) = \sum_{\mathbf{x}|\mathbf{z}} P(\boldsymbol{\xi} = \mathbf{x}) = \frac{2^{m_2(\mathbf{z})}\binom{m}{\mathbf{m}(\mathbf{z})}}{\binom{2m}{\mathbf{d}}} = \frac{2^{m_2(\mathbf{z})} m!\ \mathbf{d}!}{\mathbf{m}(\mathbf{z})!\ (2m)!}$$
$$= \frac{2^{m_2(\mathbf{z})} m! \prod_{i=1}^{n} d_i!}{(2m)! \prod_{i \le j} m_{ij}(\mathbf{z})!} \quad \text{for} \quad \mathbf{z} \in Z(\mathbf{d}) \ .$$

Consider the ordered partition $\mathbf{m}(\mathbf{z})$ and the corresponding unordered partition of m into r non-negative integers. There is a bijection between this partition and the sequence of frequencies of sites with multiplicities $0, 1, \ldots, m$ given by $\mathbf{r}(\mathbf{z}) = (r_0(\mathbf{z}), \ldots, r_m(\mathbf{z}))$ where

$$r_k(\mathbf{z}) = \sum_{i \le j} \sum I(m_{ij}(\mathbf{z}) = k) \qquad \text{for} \quad k = 0, 1, \ldots, m \ .$$

The distribution of multiplicities that is given by $\mathbf{r}(\mathbf{z})$ is called the complexity of the graph with edge sequence $\mathbf{z}$ (Frank and Shafie, 2012). It is convenient to separate frequencies of loops and non-loops and use $r(\mathbf{z}) = r_1(\mathbf{z}) + r_2(\mathbf{z})$ where

$$\mathbf{r_1}(\mathbf{z}) = (r_{10}(\mathbf{z}), \ldots, r_{1m}(\mathbf{z})) \quad \text{and} \quad \mathbf{r_2}(\mathbf{z}) = (r_{20}(\mathbf{z}), \ldots, r_{2m}(\mathbf{z}))$$

with

$$r_{1k}(\mathbf{z}) = \sum_{i=1}^{n} I(m_{ii}(\mathbf{z}) = k) \quad \text{and} \quad r_{2k}(\mathbf{z}) = \sum_{i < j} \sum I(m_{ij}(\mathbf{z}) = k) \quad \text{for} \quad k = 0, 1, \ldots, m \ .$$

Using these complexities it is possible to express the probability $P(\boldsymbol{\zeta} = \mathbf{z})$ as a function of a special summary measure of complexity according to the following:

$$P(\boldsymbol{\zeta} = \mathbf{z}) = C 2^{-t(\mathbf{z})} \ ,$$

where $C = 2^m m!\, \mathbf{d}! \,/(2m)!$ and

$$\begin{aligned} t(\mathbf{z}) &= m_1(\mathbf{z}) + \log \mathbf{m}(\mathbf{z})! \\ &= \sum_{i=1}^{n} m_{ii}(\mathbf{z}) + \sum_{i \le j} \sum \log m_{ij}(\mathbf{z})! \\ &= \sum_{k=1}^{m} k\ r_{1k}(\mathbf{z}) + \sum_{k=2}^{m} r_k(\mathbf{z}) \log k! \\ &= \sum_{k=1}^{m} (k + \log k!\,) r_{1k}(\mathbf{z}) + \sum_{k=2}^{m} r_{2k}(\mathbf{z}) \log k! \ . \end{aligned}$$

Simple graphs without loops and multiple edges have $t(\mathbf{z}) = 0$ and all simple graphs have the same probability C. More complex graphs have higher values of $t(\mathbf{z})$ and smaller probabilities. All graphs with a fixed value $t(\mathbf{z}) = t$ of the complexity measure have the same probability. The set $Z(\mathbf{d})$ of edge sequences is partitioned according to values of the complexity measure, and the set of edge sequences with complexity t is denoted

$$Z(\mathbf{d}, t) = \{\mathbf{z} \in Z(\mathbf{d}) : t(\mathbf{z}) = t\} \ .$$

The number of sequences in this set is denoted $|Z(\mathbf{d}, t)| = K(\mathbf{d}, t)$, or simply K_t if $\mathbf{d}$ is clear from context. The probability of complexity value t is given by

$$P(t(\boldsymbol{\zeta}) = t) = \sum_{\mathbf{z}|t(\mathbf{z})=t} P(\boldsymbol{\zeta} = \mathbf{z}) = CK_t 2^{-t} \ ,$$

and the conditional distribution of $\boldsymbol{\zeta}$ given complexity t is equal to

$$P(\boldsymbol{\zeta} = \mathbf{z}|t(\boldsymbol{\zeta}) = t) = \frac{C2^{-t}}{CK_t 2^{-t}} = \frac{1}{K_t}$$

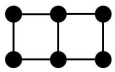
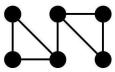
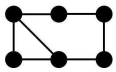
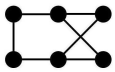
which is uniform on K_t outcomes in $Z(\mathbf{d}, t)$. Neither K_t nor the probability $P(t(\boldsymbol{\zeta}) = t)$ are monotone as functions of t. This will follow by an examination of K_t in a numerical example considered below.

Example.
Consider undirected graphs with degree sequence $\mathbf{d} = (3, 3, 2, 2, 2, 2)$. There are in total 784 possible vertex labeled graphs with edge sequences $\mathbf{z}$ in $Z(\mathbf{d})$. Table 1 lists the probability $P(t(\boldsymbol{\zeta}) = t)$, the number of vertex labeled graphs K_t, and the probability per graph $P(t(\boldsymbol{\zeta}) = t)/K_t$ for each complexity value 2^t. As seen, neither K_t nor the probability $P(t(\boldsymbol{\zeta}) = t)$ are monotone as functions of t. It is also clear that every simple graph has a higher probability of occurring than any complex graph, and that the more complex a graph is, the smaller probability it has. Note however that this does not mean that all unlabeled graphs of given complexity have the same probability of occurring. This is clarified by looking at all unlabeled simple graphs in this example. In Table 2 the number of isomophic graphs (the number of vertex labeled graphs) are listed for the four possible unlabeled simple graphs. We see in Table 2 that the third unlabeled graph has more edge sequences generating it than the others, and therefore it has the highest probability of occurring.

Table 1: Complexity distribution of graphs with degree sequence (3, 3, 2, 2, 2, 2).

Complexity (2^t)	Probability of complexity	Number of graphs	Probability per graph
1	0.230170	54	0.004262
2	0.387880	182	0.002131
4	0.252522	237	0.001065
6	0.002131	3	0.000710
8	0.098035	184	0.000533
12	0.001421	4	0.000355
16	0.024242	91	0.000266
24	0.000535	3	0.000178
32	0.002398	18	0.000133
48	0.000533	6	0.000089
64	0.000067	1	0.000067
96	0.000044	1	0.000044

Table 2: Simple graphs with degree sequence (3, 3, 2, 2, 2, 2).

Unlabeled graphs					Total
Vertex labeled graphs	12	6	24	12	54

4 Moments of Edge Multiplicities

In order to investigate the distribution of the edge multiplicites under random stub matching (RSM), we start by analyzing the moments of this distribution. The probability of coupling stubs to edges in $\boldsymbol{\xi}$ is

$$P_{ij} = P\left((\xi_{2k-1}, \xi_{2k}) = (i,j)\right) = \begin{cases} \binom{d_i}{2} / \binom{2m}{2} & \text{for } i = j \\ d_i d_j / 2m(2m-1) & \text{for } i \neq j \ , \end{cases}$$

where $\sum_{i=1}^{n}\sum_{j=1}^{n} P_{ij} = 1$. The probability of undirected edges in $\boldsymbol{\eta}$ is thus equal to

$$Q_{ij} = P\left((\eta_{2k-1}, \eta_{2k}) = (i,j)\right) = \begin{cases} P_{ii} = \binom{d_i}{2}/\binom{2m}{2} & \text{for } i = j \\ 2P_{ij} = d_i d_j/\binom{2m}{2} & \text{for } i < j \\ 0 & \text{for } i > j \ . \end{cases}$$

Note that (η_{2k-1}, η_{2k}) are identically but not independently distributed. It is convenient to introduce $Q_{ijk\ell} = P\left((\eta_{2u-1}, \eta_{2u}) = (i,j) \text{ and } (\eta_{2v-1}, \eta_{2v}) = (k,\ell)\right)$ for $u \neq v$. For $i \leq j$ and $k \leq \ell$ we have that $Q_{ijk\ell} = Q_{k\ell ij}$.

The expected values and variances of the numbers of loops and non-loops in $\boldsymbol{\eta}$ (or in the canonical edge sequence $\boldsymbol{\zeta}$) under RSM are derived by using the first and second moments of the edge multiplicities, which occasionally is shortly denoted m_{ij} when randomness is clear, i.e.

$$m_{ij} = m_{ij}(\boldsymbol{\eta}) = m_{ij}(\boldsymbol{\zeta}) = \sum_{k=1}^{m} I_{ijk},$$

where the indicators are given by

$$I_{ijk} = I\left((\eta_{2k-1}, \eta_{2k}) = (i,j)\right) = \begin{cases} 1 & \text{if } (\eta_{2k-1}, \eta_{2k}) = (i,j) \\ 0 & \text{otherwise} \ , \end{cases}$$

for $(i,j) \in R$ and $k \in E$. Now $E(I_{ijk}) = Q_{ij}$ so that

$$E(m_{ij}) = mQ_{ij} = \begin{cases} \binom{d_i}{2}/(2m-1) & \text{for } i = j \\ d_i d_j/(2m-1) & \text{for } i < j \ , \end{cases}$$

In order to obtain the variance of m_{ij} under RSM, we need the covariance between indicators I_{ijk} and $I_{ij\ell}$. They are given by

$$\mathrm{Cov}(I_{ijk}, I_{ij\ell}) = \begin{cases} Q_{ij}(1 - Q_{ij}) & \text{for } k = \ell \\ Q_{ijij} - Q_{ij}^2 & \text{for } k \neq \ell \ , \end{cases}$$

where

$$Q_{ijij} = \begin{cases} \binom{d_i}{2}\binom{d_i-2}{2}/\binom{2m}{2}\binom{2m-2}{2} & \text{for } i = j \\ d_i d_j (d_i - 1)(d_j - 1)/\binom{2m}{2}\binom{2m-2}{2} & \text{for } i < j \ . \end{cases}$$

Hence

$$\begin{aligned}
\mathrm{Var}(m_{ij}) &= \sum_{k=1}^{m}\sum_{\ell=1}^{m}\mathrm{Cov}(I_{ijk}, I_{ij\ell}) \\
&= mQ_{ij}(1-Q_{ij}) + m(m-1)(Q_{ijij} - Q_{ij}^2) \\
&= mQ_{ij}(1-mQ_{ij}) + m(m-1)Q_{ijij} \\
&= \begin{cases} \frac{\binom{d_i}{2}}{2m-1}\left(1 - \frac{\binom{d_i}{2}}{2m-1}\right) + \frac{6\binom{d_i}{4}}{(2m-1)(2m-3)} & \text{for } i = j \\ \\ \frac{d_i d_j}{(2m-1)}\left(1 - \frac{d_i d_j}{2m-1}\right) + \frac{d_i d_j (d_i-1)(d_j-1)}{(2m-1)(2m-3)} & \text{for } i < j\ . \end{cases}
\end{aligned}$$

Covariances between m_{ij} and $m_{k\ell}$ require covariances between indicators I_{iju} and $I_{k\ell v}$ for $u = 1, \ldots, m$ and $v = 1, \ldots, m$. Here $i \leq j$ and $k \leq \ell$ and, since $\mathrm{Cov}(m_{ij}, m_{k\ell}) = \mathrm{Cov}(m_{k\ell}, m_{ij})$, it is sufficient to consider $i \leq k$, and for $i = k$, only $j \leq \ell$. Explicit expressions for such covariances will be given when needed in the sequel.

The variance of m_{ij} under RSM can be written as

$$\mathrm{Var}(m_{ij}) = \sigma_{ij}^2 + \Delta_{ij} \quad \text{for } i \leq j\ .$$

Here, $\sigma_{ij}^2 = mQ_{ij}(1-Q_{ij})$ is the variance of a binomial distribution obtained by independent edge assignments (IEA) with parameters m and Q_{ij} for $i \leq j$ and

$$\Delta_{ij} = m(m-1)(Q_{ijij} - Q_{ij}^2)\ .$$

Using these expressions we can now show for which values of d_i and d_j the variance of the IEA distribution is smaller or larger than the variance of the RSM distribution of edge multiplicity. We start with the case when $i = j$ and search the sign of Δ_{ii} for values of $d_i = d$ where $2 \leq d \leq 2m-1$. By rewriting Δ_{ii} as

$$\Delta_{ii} = m(m-1)Q_{ii}\left[\frac{(d-2)(d-3)}{(2m-2)(2m-3)} - \frac{d(d-1)}{2m(2m-1)}\right]\ ,$$

and noticing that

$$\frac{d-k}{2m-k} = 1 - \frac{2m-d}{2m-k}$$

is a decreasing function of k, it follows that $\Delta_{ii} < 0$ for $d < 2m$. Thus, $\mathrm{Var}(m_{ii}) < \sigma_{ii}^2$ for $1 < d < 2m$ and $\mathrm{Var}(m_{ii}) = \sigma_{ii}^2$ only for the degenerate cases $d = 1$ and $d = 2m$ with $\sigma_{ii}^2 = 0$.

When $i < j$, set $a = \min(d_i, d_j)$ and $b = \max(d_i, d_j)$ and search the sign of Δ_{ij} for different pairs of values (a, b) with $1 \leq a \leq b$ and $a + b \leq 2m$ for $m > 1$. By rewriting Δ_{ij} as

$$\Delta_{ij} = m(m-1)Q_{ij}\left[\frac{(a-1)(b-1)}{\binom{2m-2}{2}} - \frac{ab}{\binom{2m}{2}}\right]\ ,$$

we see that Δ_{ij} has the same sign as the function

$$f(a,b) = \frac{(a-1)(b-1)}{ab} - \frac{\binom{2m-2}{2}}{\binom{2m}{2}} = \left(1-\frac{1}{a}\right)\left(1-\frac{1}{b}\right) - \left(1-\frac{1}{m}\right)\left(1-\frac{1}{m-\frac{1}{2}}\right) \ .$$

Now $f(a,b) < 0$ for $1 \le a \le b \le m-1$, and $f(1,b) < 0$ for $1 \le b \le 2m-1$. For fixed value a or fixed value b, $f(a,b)$ is increasing in the other variable. Moreover, $f(m,m) > 0$. In order to find the critical curve between positive and negative values of $f(a,b)$, we set $f(a,b) = 0$ and solve for b to get

$$b = \frac{a-1}{a\theta - 1} \ ,$$

where $\theta = (4m-3)/m(2m-1)$ and between 0 and 1. The intersection between this curve and the upper boundary $b = 2m-a$ of the (a,b)-region defined by $1 \le a \le b$ and $a+b \le 2m$ is obtained as the solution to the quadratic equation

$$a^2 - 2ma + \frac{2m-1}{\theta} = 0$$

with roots

$$a = m \pm \sqrt{\frac{m(m-1)}{4m-3}} \ .$$

The relevant root is $m - \sqrt{m(m-1)/(4m-3)}$ since $a = \min(d_i, d_j)$ cannot be larger than m. It follows that

$$f(a, 2m-a) < 0 \quad \text{for} \quad 1 \le a < m - \sqrt{\frac{m(m-1)}{4m-3}} \ ,$$

$$f(a, 2m-a) > 0 \quad \text{for} \quad m - \sqrt{\frac{m(m-1)}{4m-3}} < a \le m \ ,$$

and

$$f(a, 2m-a) = 0 \quad \text{if} \quad a = m - \sqrt{\frac{m(m-1)}{4m-3}} \quad \text{is integer.}$$

With a similar investigation of the line $b = 2m-1-a$ and the critical curve, we find no intersection and therefore $f(a,b) < 0$ for $1 \le a \le b \le 2m-1-a$. The conclusion is that

$$\Delta_{ij} > 0 \quad \text{only for} \quad m - \sqrt{\frac{m(m-1)}{4m-3}} < a = 2m-b \le m \ ,$$

that is for the $\left\lceil \sqrt{m(m-1)/(4m-3)} \right\rceil$ integer points $(a, 2m-a)$ with

$$m - \sqrt{\frac{m(m-1)}{4m-3}} < a \le m$$

on the upper boundary. Moreover, $\Delta_{ij} < 0$ for the other $m^2 - \left\lceil \sqrt{m(m-1)/(4m-3)} \right\rceil$ points (a, b) in the (a, b)-region. Thus the RSM distribution of m_{ij} has a variance that is smaller than σ_{ij}^2 unless d_i and d_j lie symmetrically around m and are given by $m \pm k$ for some non-negative integer

$$k < \sqrt{\frac{m(m-1)}{4m-3}} \ .$$

It also follows that the variance is maximal for $k = 0$ and decreases for increasing k. The case $m = 20$ is illustrated in Figure 2 where the points with positive Δ_{ij} are marked with $(*)$.

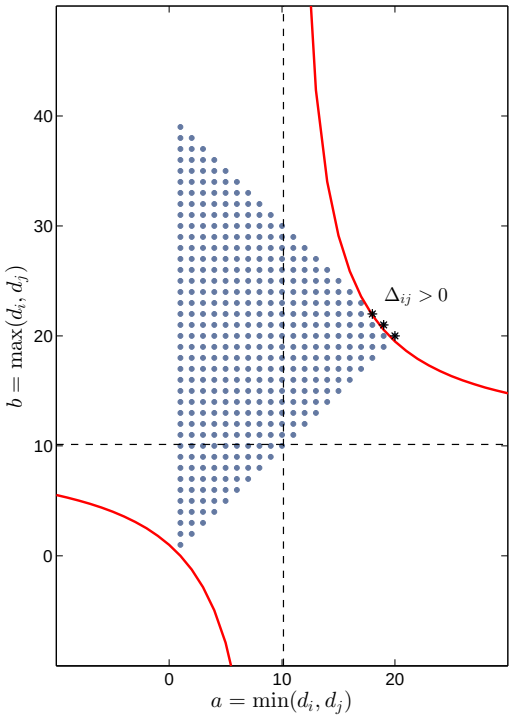


Figure 2: Points and critical curve. Points represent posssible degree pairs (a, b) at a given vertex pair in a graph with $m = 20$ edges. A critical curve is separating points with positive and negative variance difference Δ_{ij} between the edge multiplicity distributions obtained at (a, b) by random stub matching and by independent edge assignments. The points where the stub matching has larger variance are marked by $(*)$.

5 Distributions of Edge Multiplicities

In this section, some new results on distributions of edge multiplicities for graphs obtained by random stub matching (RSM) are derived. There are many asymptotic results in the literature (e.g. Bollobàs 2001) but we are also interested in exact results for fixed degree sequences. Specific results for no loops and no multiple edges have been discussed in Janson (2009) and Bollobàs (1980), where random stub matching is referred to as the configuration model or the pairing model.

The probability of no loops at vertex i, denoted P_0, is given by Janson (2009) as

$$P_0 = \prod_{j=1}^{d_i} \frac{2m - d_i - j + 1}{2m - 2j + 1} \ .$$

He gives no formula for arbitrary numbers of loops at i but notes that it is more difficult to find the probability of no multiple edges due to the complications caused by loops. We now derive a formula for an arbitrary number of loops at vertex i under RSM and get in particular a simple expression for the number of no loops at i. This technique can be generalized and used to derive the probability of arbitrary multiplicities at any site $(i, j) \in R$.

Consider the probability of v loops at vertex i under RSM denoted by $P_v = P(m_{ii}(\boldsymbol{\eta}) = v) = P(m_{ii} = v)$ for $v = 0, \ldots, m$. To find how many of the $\binom{2m}{\mathbf{d}}$ possible stub sequences that generate v loops at i, arrange m edges with v loops at i, $d_i - 2v$ edges with the remaining i-stubs, and $m - d_i + v$ other edges. This number of arrangements is given by the multinomial coefficient $\binom{m}{v,\ d_i - 2v}$. The single i-stubs have two alternative locations in the $d_i - 2v$ edges. Finally, the remaining stubs are arranged in $\binom{2m-d_i}{\mathbf{d}^*}$ ways where $\mathbf{d}^*$ is the degree sequence $\mathbf{d}$ without d_i. This leads to

$$P_v = \frac{\binom{m}{v,\ d_i - 2v} 2^{d_i - 2v} \binom{2m - d_i}{\mathbf{d}^*}}{\binom{2m}{\mathbf{d}}}$$

which simplifies to

$$P_v = \frac{\binom{m}{v,\ d_i - 2v} 2^{d_i - 2v}}{\binom{2m}{d_i}} \ .$$

In particular the probability of no loops at vertex i under RSM is equal to

$$P_0 = \frac{\binom{m}{d_i} 2^{d_i}}{\binom{2m}{d_i}} \ .$$

This formula can be developed according to the following which shows that it is equivalent to Janson's (2009) expression for P_0 as a ratio between a falling factorial from $2m - d_i$ and

a falling semifactorial from $2m-1$, both carried out for d_i factors (in fact, $d_i - 1$ factors suffice since the last one cancels):

$$\begin{aligned} P_0 &= \frac{m!\ d_i!\ (2m-d_i)!\ 2^{d_i}}{d_i!\,(m-d_i)!\ (2m)!} \\ &= \frac{m!\ 2^m\ (2m-d_i)!}{(2m)!\ (m-d_i)!\ 2^{m-d_i}} \\ &= \frac{(2m)!!\ (2m-d_i)!}{(2m)!\ (2m-2d_i)!!} \\ &= \frac{(2m-2d_i-1)!!\ (2m-d_i)!}{(2m-1)!!\ (2m-2d_i)!} \\ &= \frac{(2m-d_i)(2m-d_i-1)\cdots(2m-2d_i+1)}{(2m-1)(2m-3)\cdots(2m-2d_i+1)}\ . \end{aligned}$$

Assume that $d_i = d$ with $2 \le d \le 2m$ and consider the general probability that there are v loops at vertex i under RSM given by

$$P(m_{ii} = v) = \frac{\binom{m}{v,\ d-2v} 2^{d-2v}}{\binom{2m}{d}} \quad \text{for} \quad v = 0, 1, \ldots, \lfloor d/2 \rfloor\ .$$

This probability is also denoted P_v or $P_v(m,d)$. Set $\mathbf{P} = (P_0, \ldots, P_{\lfloor d/2 \rfloor})$. The expected value and variance of m_{ii} under RSM given in the previous section are equal to μ_{ii} and $\sigma_{ii}^2 + \Delta_{ii}$, where $\mu_{ii} = \binom{d}{2}/(2m-1)$ and $\sigma_{ii}^2 = mQ_{ii}(1-Q_{ii}) = \mu_{ii}\left(1 - \frac{\mu_{ii}}{m}\right)$ are the mean and variance of the IEA distribution $\mathbf{B} = (B_0, \ldots, B_m)$ with parameters m and Q_{ii}. The range of the multiplicity distribution under IEA is $v = 0, 1, \ldots, m$ and the range of the multiplicity distribution under RSM is smaller. Its proportion is $(\lfloor d/2 \rfloor + 1)/(m+1)$ of the range of the IEA distribution. Table 3 gives these distributions for the case $m = 10$ and $d = 10$. Also presented in Table 3 is a measure of the discrepancy between the distributions given by the information divergence

$$D(\mathbf{P}, \mathbf{B}) = \sum_{v: P_v > 0} P_v \log \frac{P_v}{B_v}\ .$$

The log-likelihood ratios can be of any sign but their weighted sum, the divergence $D(\mathbf{P}, \mathbf{B})$, is non-negative and zero only when there is no discrepancy between the two distributions (see e.g. Frank 2011). Figure 3 shows how divergence varies for different values of $d = 2, \ldots, 2m-1$ for $m = 40$, and how the divergence varies for different stub proportions $d/2m$ (or range proportions $(\lfloor d/2 \rfloor + 1)/(m+1)$) for some values of m.

Table 3: The probability distribution of loop multiplicity at a vertex of degree $d = 10$ when $m = 10$ edges are formed by random stub matching (RSM). It is compared by information divergence to a binomial distribution obtained by $m = 10$ independent edge assignments (IEA).

Number of loops	Probability under RSM	Probability under IEA	Weighted log-likelihood ratio
0	0.005542	0.067011	-0.019930
1	0.124705	0.207964	-0.092044
2	0.436468	0.290433	0.256636
3	0.363723	0.240358	0.217242
4	0.068198	0.130539	-0.063849
5	0.001364	0.048615	-0.007164
6	0	0.012573	0
7	0	0.002230	0
8	0	0.000259	0
9	0	0.000018	0
10	0	0.000001	0
			Divergence = 0.290891

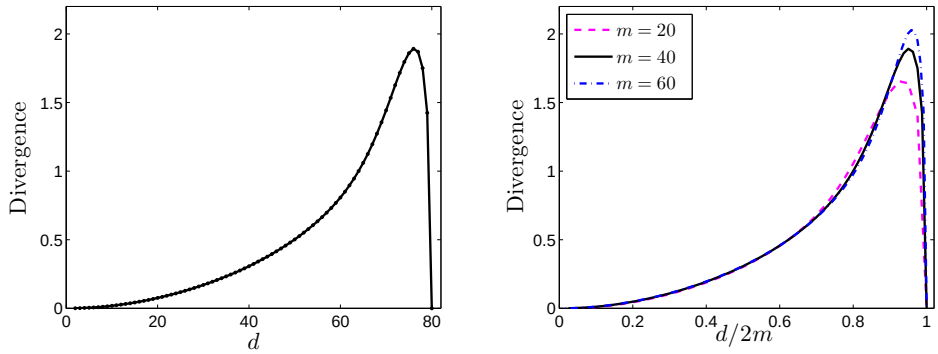


Figure 3: Information divergence between random stub matching and independent edge assignments for the distributions of loop multiplicity at a vertex of degree d in a graph with m edges. Information divergence is plotted against degree d for $m = 40$ and against stub proportion $d/2m$ for $m = 20, 40, 60$.

In order to compare the two distributions we also use the entropy which characterize flatness of the distributions $\mathbf{P}$ and $\mathbf{B}$. The entropies are equal to

$$h(\mathbf{P}) = \sum_{v:P_v>0} -P_v \log P_v$$

and

$$h(\mathbf{B}) = \sum_{v:B_v>0} -B_v \log B_v \ .$$

Upper bounds to the entropies are given by their logarithmic ranges:

$$h(\mathbf{P}) \leq \log\left(\lfloor d/2 \rfloor + 1\right)$$

and

$$h(\mathbf{B}) \leq \log(m+1) \ .$$

For the case where $m = 10$ and $d = 10$ we have that $h(\mathbf{P}) = 1.746$ and $h(\mathbf{B}) = 2.443$. Figure 4 shows how entropies vary for different stub proportions $d/2m$ for $m = 20, 40, 60$. We see that stub matching has lower entropy and is more symmetric around 0.5 than the entropy corresponding to independent edge assignments. The latter entropy is skew to the right and has its maximum when loop probability $\binom{d}{2}/\binom{2m}{2}$ is about $1/2$ which occurs for the stub proportion close to $1/\sqrt{2} \approx 0.7$.

The asymptotic entropies (Frank and Nowicki 1989) of the loop multiplicity distribution under RSM and under IEA are obtained by normal approximations given by

$$h(\mathbf{P}) \approx \frac{1}{2} \log\left[2\pi e(\sigma_{ii}^2 + \Delta_{ii})\right] \ ,$$

and

$$h(\mathbf{B}) \approx \frac{1}{2} \log\left[2\pi e \sigma_{ii}^2\right] \ .$$

For the case where $m = 10$ and $d = 10$, these approximations are equal to $h(\mathbf{P}) \approx 1.747$ and $h(\mathbf{B}) \approx 2.474$. Figure 5 shows how these entropy approximations (dotted lines) vary for the same cases as in Figure 4, i.e. for different stub proportions $d/2m$ for $m = 20, 40, 60$. We see that the approximations are close to their true values for all cases shown.

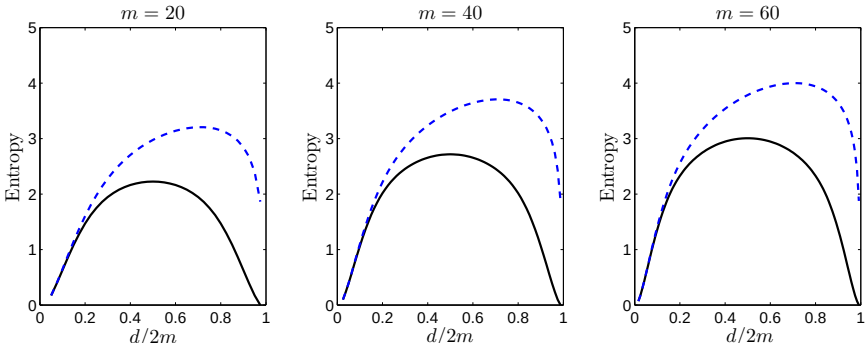


Figure 4: Entropy of the distribution of loop multiplicity under random stub matching (solid lines) and independent edge assignments (dashed lines) at a vertex of degree d in a graph with m edges. Entropy is plotted against stub proportion $d/2m$ for $m = 20, 40, 60$.

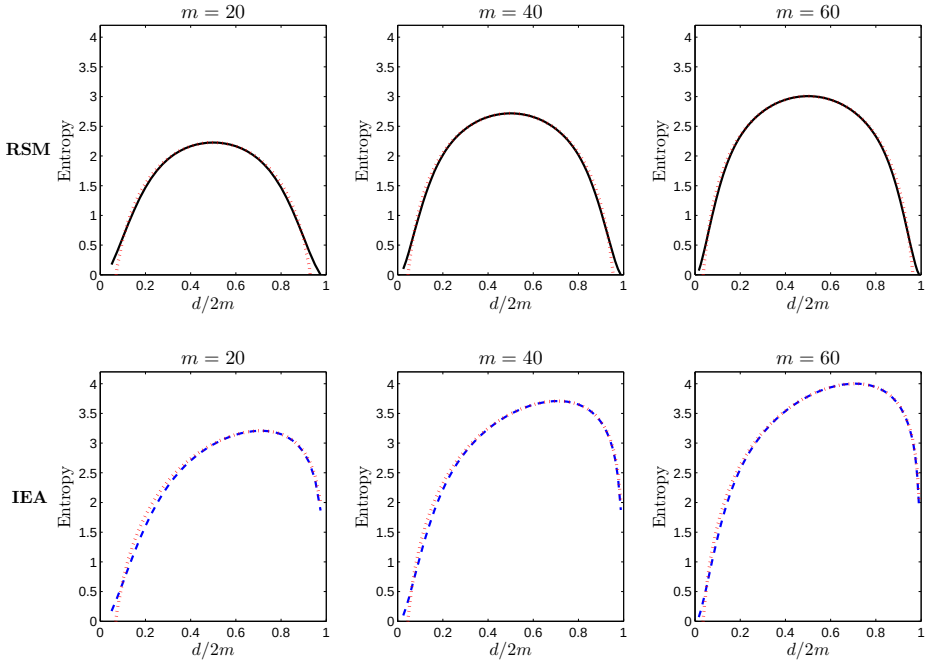


Figure 5: Entropy of the distribution of loop multiplicity under random stub matching (solid lines) and independent edge assignments (dashed lines) at a vertex of degree d in a graph with m edges. The entropy approximations are illustrated with dotted lines for all cases. Entropy is plotted against stub proportion $d/2m$ for $m = 20, 40, 60$.

We now turn to edge multiplicities $m_{ij} = m_{ij}(\boldsymbol{\eta})$ for $i < j$ under RSM, and start with the joint probability distribution of the multiplicities $(m_{ii}, m_{jj}, m_{ij}) = (m_{ii}(\boldsymbol{\eta}), m_{jj}(\boldsymbol{\eta}), m_{ij}(\boldsymbol{\eta}))$ which is denoted $P_{uvw} = P((m_{ii}, m_{jj}, m_{ij})) = (u, v, w))$. Applying a similar argument as before for i-loops, j-loops, (i, j)-edges, remaining i-stubs and j-stubs, we obtain after simplification the following formula for the trivariate probabilities under RSM:

$$P_{uvw} = \frac{\binom{m}{u,\ v,\ w,\ d_i-2u-w,\ d_j-2v-w} 2^{d_i+d_j-2u-2v-w}}{\binom{2m}{d_i,\ d_j}} .$$

To specify possible outcomes of (u, v, w) let a and b denote the smallest and largest number of stubs at vertices i and j, and c denote the number of stubs at other vertices, i.e. $c = 2m-a-b$. Table 4 gives the number of stubs of each category occurring at the edges within and between categories. There are u loops with $2u$ stubs at the vertex with a stubs, v loops with $2v$ stubs at the vertex with b stubs, and w edges with w stubs of each kind between these two vertices. There are $a-2u-w$ and $b-2v-w$ remaining stubs at these two vertices, and they combine to edges with the same number of other stubs. Since there is a total of c other stubs, there remain $c-(a-2u-w)-(b-2v-w) = 2(m-a-b+u+v+w)$ other stubs giving room for $m-a-b+u+v+w$ other loops or edges. Note that $2u+w \leq a$, $2v+w \leq b$ and $(2u + w) + (2v + w) \geq a + b - c$ are required to achieve non-negative frequencies.

Table 4: Number of stubs of each vertex category at edges or loops within and between categories.

	Vertex i	Vertex j	Other vertices	Total
Vertex i	$2u$	w	$a - 2u - w$	a
Vertex j	w	$2v$	$b - 2v - w$	b
Other vertices	$a - 2u - w$	$b - 2v - w$	$2(m - a - b + u + v + w)$	c
Total	a	b	c	$2m$

The set of possible outcomes (u, v, w) is illustrated in Figure 6 and correspond to marked points in the shaded regions with possible stub frequencies $(2u + w, 2v + w)$ of the same parity. Four different cases are numerically illustrated in Figure 6 where $(a, b) = (3, 7)$ and the value of c is varied and given by the vertical distance from $a+b$ to the upper point of the digonal. First, consider $b \leq a+b-c \leq a+b$ which corresponds to $0 \leq c \leq a$ and choose $c = 2$ to illustrate. There are four possible points in the triangular region, namely $(1, 7)$, $(2, 6)$,

$(3,5)$, $(3,7)$ corresponding to (u,v,w) equal to $(0,3,1)$, $(1,3,0)$, $(0,2,2)$, $(1,2,1)$, $(0,1,3)$, $(1,3,1)$, $(0,2,3)$. These (u,v,w) are obtained by choosing $w = 0,2,\ldots$ or $w = 1,3,\ldots$ so that $2u$ and $2v$ get even. It can be shown that the possible point $(2u+w, 2v+w)$ in the shaded region corresponds to

$$1 + \left\lfloor \frac{\min(2u+w, 2v+w)}{2} \right\rfloor$$

possible outcomes (u,v,w). Second, consider the case where $a \leq a+b-c \leq b$ which corresponds to $a \leq c \leq b$ and choose $c = 6$. Third, consider $0 \leq a+b-c \leq a$ which corresponds to $a \leq b \leq c \leq a+b$ and choose $c = 8$, and finally, fourth consider $a+b-c \leq 0$ which corresponds to $a+b \leq 0$ and choose $c = 12$ to illustrate the case with maximal number of outcomes.

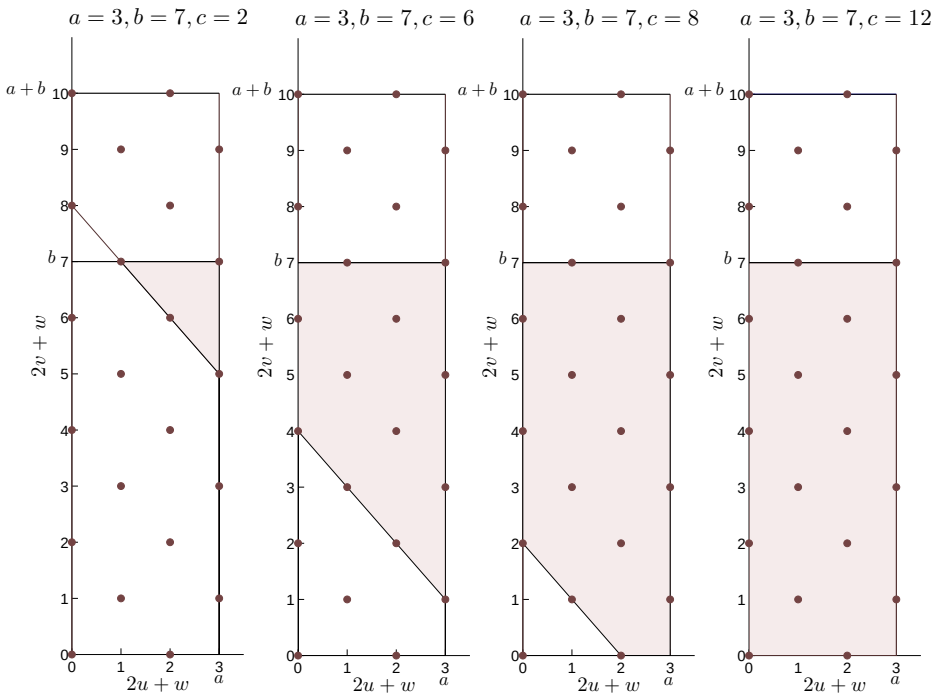


Figure 6: The possible outcomes of (u,v,w) in Table 4 correspond to the shaded region with stub frequencies $(2u+w, 2v+w)$ of the same parity. The four cases illustrate how the region varies when the number of other stubs c is smaller than a, between a and b, between b and $a+b$, and larger than $a+b$.

Letting $u \le \lfloor (a-w)/2 \rfloor = \alpha_w$, $v \le \lfloor (b-w)/2 \rfloor = \beta_w$, and $u+v \ge a+b-m-w = \gamma_w$, the total number of possible outcomes of (u,v,w) denoted K is given by

$$K = \sum_{w=0}^{a} K_w \ ,$$

where

$$K_w = \begin{cases} \binom{\alpha_w+\beta_w-\gamma_w+2}{2} & \text{if } \gamma_w \ge \beta_w \\ (\alpha_w+1)(\beta_w-\gamma_w) + \binom{\alpha_w+2}{2} & \text{if } \alpha_w \le \gamma_w \le \beta_w \\ (\alpha_w+1)(\beta_w+1) - \binom{\gamma_w+1}{2} & \text{if } \gamma_w \le \alpha_w \ , \end{cases}$$

providing an upper bound to the entropy. Table 5 gives a numerical example of this result for the second case in Figure 6, i.e. when $a=3$, $b=7$ and $c=6$. The twelve points in Figure 6 for this case can be individually checked for possible (u,v,w). There are six points with two outcomes and six points with one outcome of (u,v,w), thus a total of 18 outcomes. Using the formula for K_w we also find that the total number of possible outcomes here is given by

$$K = \sum_{w=0}^{3} K_w = 18 \ ,$$

implying that the entropy of this distribution is

$$h(\mathbf{P}) = \sum_{uvw:P_{uvw}>0} -P_{uvw} \log P_{uvw} \le \log(18) = 4.170 \ ,$$

where $\mathbf{P} = (P_{uvw} : \text{ all possible } (u,v,w))$. The exact entropy in this case turns out to be 3.525. Further comparisons are presented in Table 6.

Table 5: The total number of possible outcomes of (u,v,w) in Table 4 when $a=3$, $b=7$ and $c=6$, where $u \le \alpha_w$, $v \le \beta_w$, and $u+v \ge \gamma_w$.

w	α_w	β_w	γ_w	K_w
0	1	3	2	5
1	1	3	1	7
2	0	2	0	3
3	0	2	-1	3
				$K=18$

If the edges are assumed to be independently assigned to sites, the IEA distribution for (m_{ii}, m_{jj}, m_{ij}) is multinomial distribution with parameters m and $[Q_{ii}\ Q_{jj}\ Q_{ij}\ (1 - Q_{ii} - Q_{jj} - Q_{ij})]$ where Q_{ij} for $i \leq j$ is defined as earlier. This distribution is shortly denoted $\mathbf{B}$ with probabilities B_{uvw} for $\binom{m+3}{3}$ different outcomes $(u, v, w, m-u-v-w)$ that are ordered partitions of m into four non-negative integers. Thus, the entropy of this distribution is

$$h(\mathbf{B}) = \sum_{uvw:B_{uvw}>0} -B_{uvw} \log B_{uvw} \leq \log \binom{m+3}{3}$$

and using the normal approximation to the multinomial distribution we obtain the approximate entropy

$$h(\mathbf{B}) \approx \frac{1}{2} \log \left[(2\pi e)^3\ \det(\mathbf{\Sigma}_{\text{IEA}}) \right] \ ,$$

where $\mathbf{\Sigma}_{\text{IEA}}$ is the covariance matrix of (m_{ii}, m_{jj}, m_{ij}) under IEA given by

$$\mathbf{\Sigma}_{\text{IEA}} = m \begin{pmatrix} Q_{ii}(1-Q_{ii}) & -Q_{ii}Q_{jj} & -Q_{ii}Q_{ij} \\ -Q_{jj}Q_{ii} & Q_{jj}(1-Q_{jj}) & -Q_{jj}Q_{ij} \\ -Q_{ij}Q_{ii} & -Q_{ij}Q_{jj} & Q_{ij}(1-Q_{ij}) \end{pmatrix} .$$

The determinant of $\mathbf{\Sigma}_{\text{IEA}}$ is given by

$$\begin{aligned} \det(\mathbf{\Sigma}_{\text{IEA}}) &= m^3 Q_{ii}Q_{jj}Q_{ij}[(1-Q_{ii})(1-Q_{jj})(1-Q_{ij}) - 2Q_{ii}Q_{jj}Q_{ij} \\ &\qquad - (1-Q_{ii})Q_{jj}Q_{ij} - (1-Q_{jj})Q_{ii}Q_{ij} - (1-Q_{ij})Q_{ii}Q_{jj}] \\ &= m^3 Q_{ii}Q_{jj}Q_{ij}(1 - Q_{ii} - Q_{jj} - Q_{ij}) \ , \end{aligned}$$

so that

$$h(\mathbf{B}) \approx \frac{1}{2} \log \left[(2\pi em)^3 Q_{ii}Q_{jj}Q_{ij}(1 - Q_{ii} - Q_{jj} - Q_{ij}) \right] \ .$$

The approximate entropy of the distribution of (m_{ii}, m_{jj}, m_{ij}) under RSM is

$$h(\mathbf{P}) \approx \frac{1}{2} \log \left[(2\pi e)^3\ \det(\mathbf{\Sigma}_{\text{RSM}})\ \right] \ ,$$

where $\det(\mathbf{\Sigma}_{\text{RSM}})$ is the determinant of the covariance matrix. The covariance matrix of (m_{ii}, m_{jj}, m_{ij}) under RSM is given by

$$\mathbf{\Sigma}_{\text{RSM}} = \mathbf{\Sigma}_{\text{IEA}} + \mathbf{\Delta} \ ,$$

where

$$\mathbf{\Delta} = m(m-1) \begin{pmatrix} Q_{iiii} - Q_{ii}^2 & Q_{iijj} - Q_{ii}Q_{jj} & Q_{iiij} - Q_{ii}Q_{ij} \\ Q_{iijj} - Q_{ii}Q_{jj} & Q_{jjjj} - Q_{jj}^2 & Q_{ijjj} - Q_{ij}Q_{jj} \\ Q_{iiij} - Q_{ii}Q_{ij} & Q_{ijjj} - Q_{ij}Q_{jj} & Q_{ijij} - Q_{ij}^2 \end{pmatrix} \ ,$$

with

$$Q_{iiii} = \frac{\binom{d_i}{2}\binom{d_i-2}{2}}{\binom{2m}{2}\binom{2m-2}{2}} \ , \quad Q_{iijj} = \frac{\binom{d_i}{2}\binom{d_j}{2}}{\binom{2m}{2}\binom{2m-2}{2}} \ ,$$

$$Q_{iiij} = \frac{\binom{d_i}{2}(d_i-2)d_j}{\binom{2m}{2}\binom{2m-2}{2}} \ , \quad Q_{ijij} = \frac{d_i(d_i-1)d_j(d_j-1)}{\binom{2m}{2}\binom{2m-2}{2}} \ ,$$

and note that $Q_{ijk\ell} = Q_{k\ell ij}$ for all $i \le j$ and $k \le \ell$.

Table 6 illustrates the different entropies presented here for some cases with $a = 3$, $b = 7$ and where the total edge frequency m is varied, including the four cases given in Figure 6. Also presented in this table is the information divergence between the RSM and IEA distribution of (m_{ii}, m_{jj}, m_{ij}) :

$$D(\mathbf{P}, \mathbf{B}) = \sum_{uvw:P_{uvw}>0} P_{uvw} \log \frac{P_{uvw}}{B_{uvw}} \ .$$

We see in Table 6 that the approximate entropies are close to the entropies of both the IEA and RSM distributions indicating that both distributions are fairly well approximated by the normal distribution. We also note that as m increases, the RSM entropy moves towards that of the IEA. This can also be seen by the divergence values which are decreasing towards zero for increasing m. The common limiting distributions of RSM and IEA is a one-point distribution at $(u, v, w) = (0, 0, 0)$ so the exact entropies tend to zero. The upper bounds for the RSM distributions tend to $\log(22)$ since there are 22 points (u, v, w) corresponding to the last case shown in Figure 6. This limit is achieved already for $m = 10$.

The distribution of a single non-loop multiplicity $m_{ij} = m_{ij}(\boldsymbol{\eta})$ for $i < j$ under RSM is given as a marginal in the trivariate distribution of (m_{ii}, m_{jj}, m_{ij}). It is obtained by summing over the numbers of loops at vertices i and j. Thus,

$$P(m_{ij} = w) = P_{\cdot\cdot w} = \sum_{u=0}^{\lfloor \frac{a-w}{2} \rfloor} \sum_{v=0}^{\lfloor \frac{b-w}{2} \rfloor} P_{uvw} \ , \quad \text{for } w = 0, 1, \ldots, a \ ,$$

where $a = \min(d_i, d_j) \ge 1$ and $b = \max(d_i, d_j)$ with $a + b \le 2m$. Note that not all $P_{uvw} > 0$. For the special case when $n = 2$, $a + b = 2m$ and $u + v + w = m$, we get $a = 2u - w$ and $b = 2v - w$ and the marginal distribution of m_{12} simplifies to

$$P(m_{12} = w) = P_w = \frac{\binom{m}{u,\ v,\ w} 2^w}{\binom{2m}{a}} = \frac{m!\ 2^w\ a!\ (2m-a)!}{(2m)!\ (\frac{a-w}{2})!\ (\frac{b-w}{2})!\ w!} \ ,$$

where $w = 0, 2, \ldots, a$ if a and b are even, and $w = 1, 3, \ldots, a$ if a and b are odd. For this case we cannot expect the binomial distribution under IEA to be an adequate approximation. Obviously the IEA distribution $\mathbf{B} = (B_w : w = 0, 1, \ldots, m)$ with parameters m and

Table 6: Entropy of the joint edge multiplicity distribution under random stub matching (RSM), independent edge assignments (IEA) and the entropy approximations for these distributions where number of stubs are $a = 3$, $b = 7$ and the total edge frequency is $m = 6, 8, 9, 11, 20, 30, 40, 50, 60$, thus including the four cases shown in Figure 6. Also given is the divergence between these two distributions.

	Entropy RSM			Entropy IEA			
m	Upper bound	Exact	Approximate	Upper bound	Exact	Approximate	Divergence
6	2.81	2.36	2.31	6.39	5.07	5.31	1.88
8	4.17	3.53	3.61	7.37	4.87	5.13	0.76
9	4.39	3.64	3.70	7.78	4.68	4.94	0.55
11	4.46	3.61	3.65	8.51	4.30	4.58	0.34
20	4.46	2.96	2.90	10.79	3.17	3.36	0.10
30	4.46	2.40	2.19	12.41	2.50	2.48	0.04
40	4.46	2.03	1.65	13.59	2.08	1.86	0.02
50	4.46	1.77	1.21	14.52	1.80	1.38	0.02
60	4.46	1.57	0.84	15.28	1.59	0.10	0.01

$ab/\binom{2m}{2} = a(2m-a)/m(2m-1)$ gives positive probabilities to all outcomes whereas the RSM distribution of edge multiplicity, $\mathbf{P} = (P_w : w = 0, 1, \ldots, m)$, has zero probabilities for all even or all odd outcomes. According to the results in Section 4, it is only for this special case, $n = 2$, that the RSM distribution can have a variance $\sigma_{ij}^2 + \Delta_{ij}$ that is larger than the variance σ_{ij}^2 of the IEA distribution. This occurs when a and b lie at the same distance from m, and this distance is strictly less than $\sqrt{m(m-1)/(4m-3)}$. Thus $\Delta_{ij} > 0$ for only one choice $(a, 2m-a) = (m, m)$ if $m < 5$, two choices (m, m) and $(m-1, m+1)$ if $5 \leq m < 17$, three choices if $17 \leq m < 37$, four choices if $37 \leq m < 65$, five choices if $65 \leq m < 101$, and so forth. Of the m cases of $(a, 2m-a)$ only $\left\lceil \sqrt{m(m-1)/(4m-3)} \right\rceil$ have a variance larger than σ_{ij}^2, so even if the number of cases increases with increasing m, the proportion of cases decreases towards zero. This is illustrated in Figure 7, where we also notice that the proportion is not monotonically decreasing.

Table 7 gives the RSM distribution of edge multiplicity and the corresponding IEA distribution for the case $m = 10$ and $(a, 2m-a) = (10, 10)$. Also presented in Table 7 is the information divergence between these distributions. The entropies for this example are equal to $h(\mathbf{P}) = 1.746$ and $h(\mathbf{B}) = 2.704$. The asymptotic entropies for the edge multiplicity

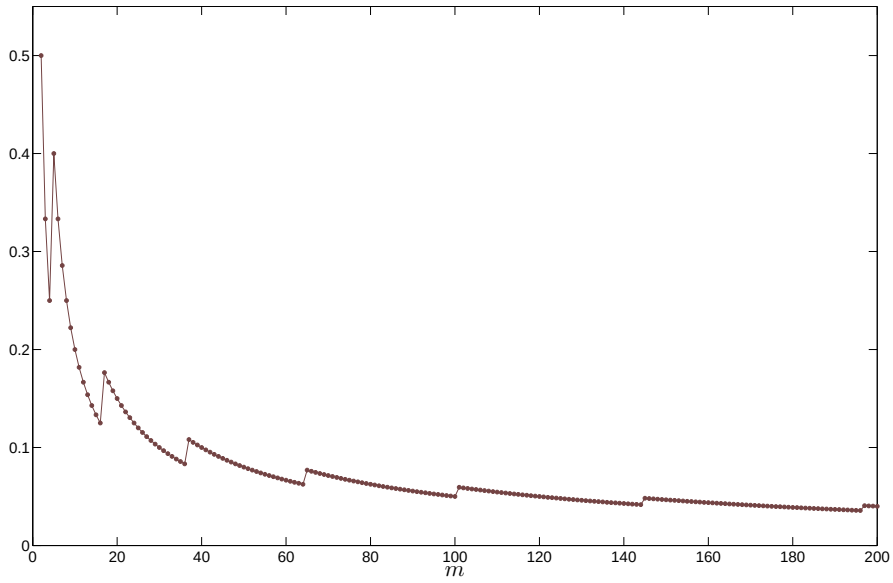


Figure 7: Proportion of the degree pairs $(a, 2m-a)$ for $a = 1, \ldots, m$ with edge multiplicity variance larger for random stub matching than for independent edge assignments. The proportions are plotted against edge frequency m.

distributions under RSM and IEA give the following approximations:

$$h(\mathbf{P}) \approx \frac{1}{2} \log \left[2\pi e(\sigma_{ij}^2 + \Delta_{ij}) \right]$$

and

$$h(\mathbf{B}) \approx \frac{1}{2} \log \left[2\pi e \sigma_{ij}^2 \right] \ .$$

For the case where $m = 10$ and $(a, 2m-a) = (10, 10)$, these approximations are equal to $h(\mathbf{P}) \approx 2.747$ and $h(\mathbf{B}) \approx 2.706$.

Figure 8 shows divergence for m_{ij} at degree pairs $(a, 2m-a)$ for different a when $m = 10$, and how it varies for different proportions a/m for some selected values of m. Figure 9 highlights how $h(\mathbf{P})$ and $h(\mathbf{B})$ vary for different a when m and Figure 10 compares the entropy approximations to their true values. The deviation between entropy values in Figure 10 indicates that edge multiplicity under RSM is poorly approximated by a normal distribution. However, the approximate entropies are close to the true values under IEA.

Table 7: The probability distribution of edge multiplicity at a pair of vertices with degree pair $(a, 2m - a) = (10, 10)$ when $m = 10$ edges are formed by random stub matching (RSM). It is compared by information divergence to the binomial distribution obtained by independent edge assignments (IEA).

Number of edges	Probability under RSM	Probability under IEA	Weighted log-likelihood ratio
0	0.001364	0.000569	0.001721
1	0	0.006319	0
2	0.068198	0.031595	0.075702
3	0	0.093614	0
4	0.363723	0.182028	0.363242
5	0	0.242704	0
6	0.436468	0.224726	0.418008
7	0	0.142683	0
8	0.1247050	0.059451	0.133277
9	0	0.014679	0
10	0.0055424	0.001631	0.009781
			Divergence = 1.001733

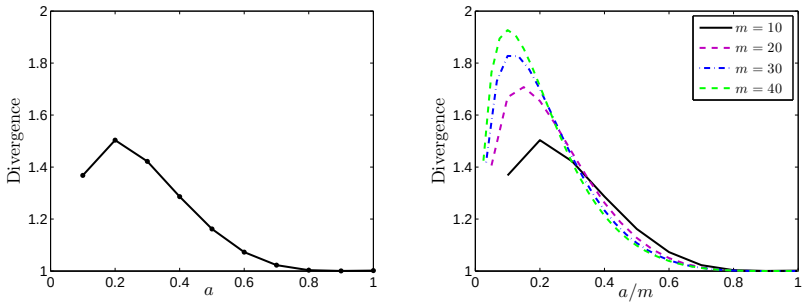


Figure 8: Information divergence between random stub matching and independent edge assignments for the distributions of edge multiplicity at a pair of vertices with degree pair $(a, 2m - a)$ in a graph with m edges. Information divergence is plotted against different degrees a for $m = 10$ and against different proportions a/m for $m = 10, 20, 30, 40$.

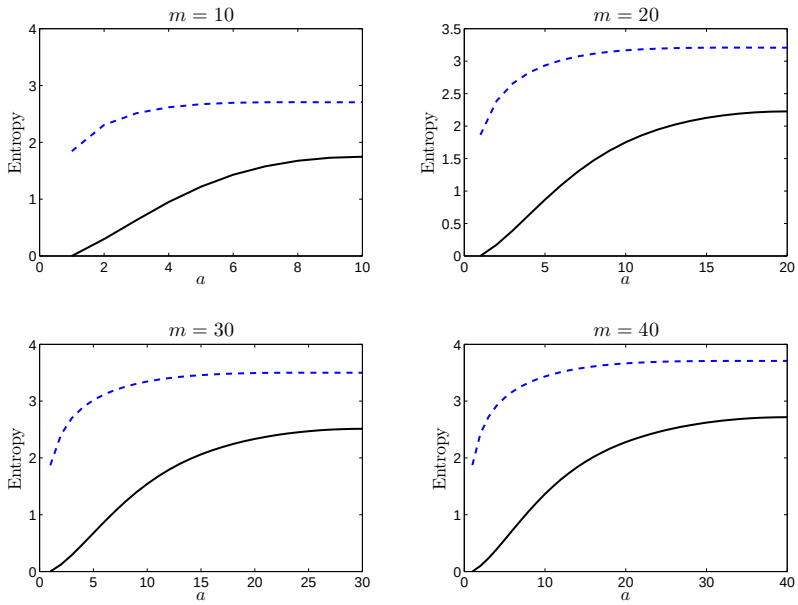


Figure 9: Entropy of the distribution of edge multiplicity under random stub matching (solid lines) and independent edge assignments (dashed lines) at a pair of vertices with degree pair $(a, 2m - a)$ in a graph with m edges. Entropy is plotted against different degrees a for $m = 10, 20, 30, 40$.

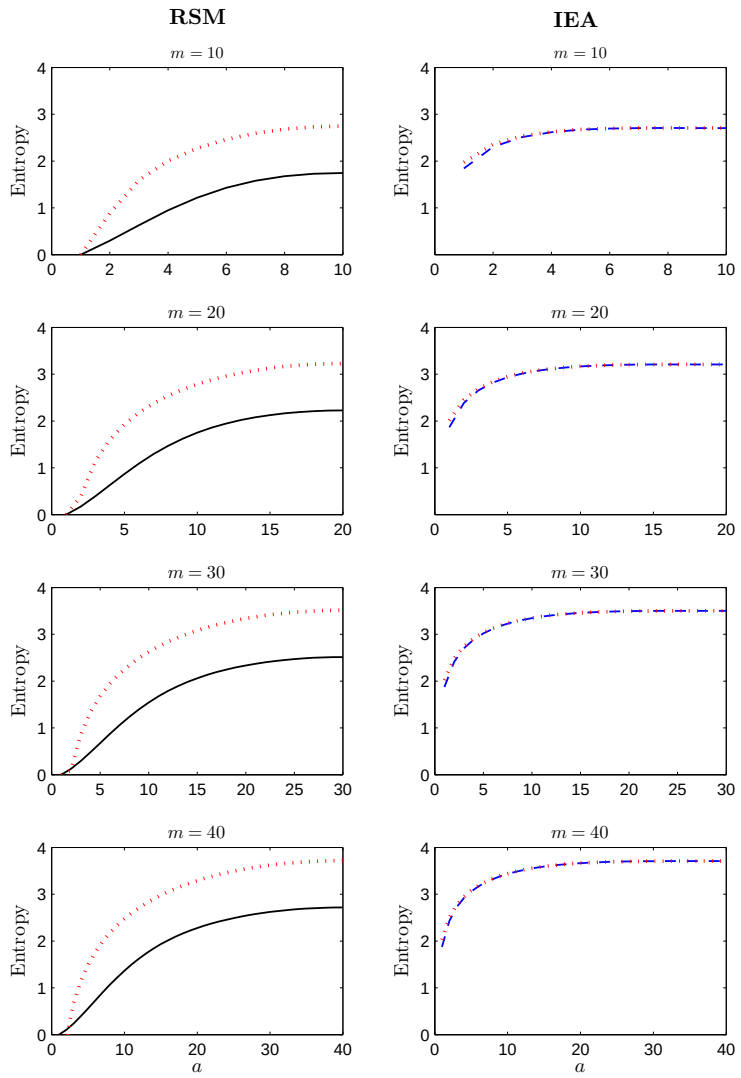


Figure 10: Entropy of the distribution of edge multiplicity under random stub matching (solid lines) and independent edge assignments (dashed lines) at a pair of vertices with degree pair $(a, 2m - a)$ in a graph with m edges. The entropy approximations are illustrated with dotted lines for all cases. Entropy is plotted against different degrees a for $m = 10, 20, 30, 40$.

This section is concluded with an illustration of how the divergence between the probability distributions of edge frequency m_{ij} for $i < j$ under RSM and under IEA vary for different numbers of stubs at vertex i and vertex j, i.e. for different ordered degree pairs (a, b) with $1 \leq a \leq b$ and $a + b \leq 2m$ where m is the total edge frequency. The case $m = 15$ is illustrated in Figure 11 where divergence $D(\mathbf{P}, \mathbf{B})$ is plotted against degree pairs (a, b) using a color coding of standardized divergence values applied to the unit squares located around points (a, b). The divergences for all possible degree pairs (a, b) are calculated and their maxima are determined. Standardized divergence values are obtained by dividing with the maxima. Every 10th percentile of this standardized distribution is then calculated and assigned a color where darker colors represent higher divergences, i.e. darker colors are assigned to unit squares where the RSM distribution deviates the most from the IEA distribution. Letting $c = 2m - a - b$ denote the number of stubs at other vertices than i and j, border lines are drawn in Figure 11 where c is equal to the stub frequencies a and b. These two border lines $b = 2m - 2a$ and $b = m - a/2$, together with the border lines $b = 2m - a$ and $b = a$, divide the figure in three regions corresponding to whether c is smaller than, or larger than, or between the two stub frequencies a and b. The upper region in Figure 11 represents cases where $c \leq a \leq b$. Here, we have the majority of the high divergence values implying that the RSM distribution and the IEA distribution deviates the most. The middle region in Figure 11 represents cases where $a \leq c \leq b$. Here, the majority of the region has a brighter color implying less deviation between the RSM distribution and the IEA distribution. The same applies for the lower region in Figure 11 which represents cases where $a \leq b \leq c$. Here, even less deviation is seen between the two distributions. Thus we can conclude that the more stubs we have at other vertices than i and j, the more resemblance we have between the distributions of m_{ij} under RSM and IEA.

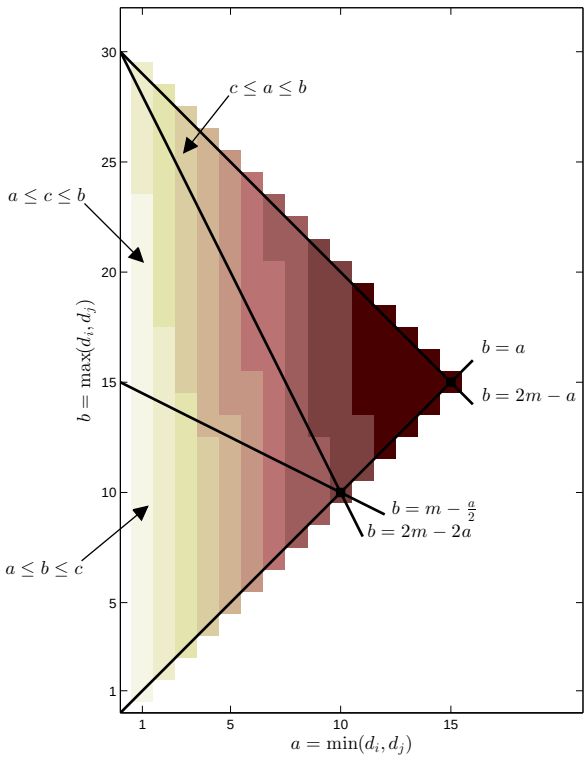


Figure 11: Divergence between the multiplicity distribution under random stub matching and under independent edge assignments for edges between two vertices with ordered degree pair (a, b) when total edge frequency is $m = 15$ and total number of stubs is $2m = a + b + c$. A darker color at the unit squares located around points (a, b) represent a larger divergence than a brighter color.

6 Distributions of Multigraphs

So far we have been considering the local structure of multigraphs by examining the distribution of edge multiplicities $m_{ij}(\boldsymbol{\eta})$ for $i \leq j$. We now turn to the global structure by studying the distribution of multigraphs, i.e. the distribution of $\mathbf{m}(\boldsymbol{\eta})$. Entropy measures are used to compare the distributions of multigraphs under RSM and under IEA.

In Section 3, the probability of generating undirected multigraphs under RSM was shown to be given by

$$P(\mathbf{m}(\boldsymbol{\eta}) = \mathbf{m}) = P_{\mathbf{m}}(\mathbf{d}) = \frac{2^{m_2}\binom{m}{\mathbf{m}}}{\binom{2m}{\mathbf{d}}} \ .$$

This probability depends on a single statistic, the complexity measure $t(\boldsymbol{\eta}) = m_1(\boldsymbol{\eta}) + \log \mathbf{m}(\boldsymbol{\eta})!$, and can be written as

$$P_{\mathbf{m}}(\mathbf{d}) = C 2^{-t} \ ,$$

where $t = m_1 + \log \mathbf{m}!$ is the outcome of $t(\boldsymbol{\eta})$ and $C = 2^m m!\, \mathbf{d}! \,/(2m)!$ is a constant. Letting $\mathbf{P} = (P_{\mathbf{m}}(\mathbf{d}) : \text{ all different } \mathbf{m})$ denote the probability distribution of multigraphs under RSM, its entropy $H(\mathbf{m}(\boldsymbol{\eta})) = h(\mathbf{P})$ is given by

$$\begin{aligned} h(\mathbf{P}) &= \sum_{\mathbf{m}: P_{\mathbf{m}} > 0} -P_{\mathbf{m}} \log P_{\mathbf{m}} \\ &= E(t(\boldsymbol{\eta})) - \log C \ , \end{aligned}$$

where the expected value of $t(\boldsymbol{\eta})$ is

$$E(t(\boldsymbol{\eta})) = E(m_1) + E(\log \mathbf{m}!) \ .$$

The first term of the above sum is the expected number of loops $m_1 = m_1(\boldsymbol{\eta})$ under RSM which by using the results in Section 4 is obtained as

$$E(m_1) = \sum_{i=1}^{n} E(m_{ii}) = m \sum_{i=1}^{n} Q_{ii} = \frac{1}{2m-1} \sum_{i=1}^{n} \binom{d_i}{2} \ .$$

The second term of the expression for the expected value of $t(\boldsymbol{\eta})$ can be expanded to

$$E(\log \mathbf{m}!) = \sum_{k=2}^{m} \log k! \sum_{i \leq j} \sum P(m_{ij} = k) \ ,$$

where the probabilities need to be considered for vertex pairs with different degree pairs. Letting $P(m_{ii} = k) = P_k(m, a)$ for $d_i = a$, $P(m_{ij} = k) = P_k(m, a, b)$ for $a = \min(d_i, d_j)$ and $b = \max(d_i, d_j)$, $\sum_{i=1}^{n} I(d_i = a) = n_a$, $\sum_{i=1}^{n} I(d_i = b) = n_b$ and

$$\sum_{i<j} \sum I\left(\min(d_i, d_j) = a, \max(d_i, d_j) = b\right) = \begin{cases} n_a n_b & \text{for } a < b \\ \binom{n_a}{2} & \text{for } a = b \ , \end{cases}$$

we have that

$$\sum_{i\le j}\sum P(m_{ij}=k) = \sum_a n_a P_k(m,a) + \sum_a \binom{n_a}{2} P_k(m,a,a) + \sum_{a<b}\sum n_a n_b P_k(m,a,b) \ .$$

These expansions yield the formula for the exact entropy. In particular, for regular graphs with the same degree $d = 2m/n$ at each vertex, the expected number of loops under RSM is given by

$$E(m_1) = \frac{n}{(2m-1)}\binom{d}{2} = \frac{nd(d-1)}{2(nd-1)} \ ,$$

and the exact entropy is simplified to

$$h(\mathbf{P}) = \frac{nd(d-1)}{2(nd-1)} + \sum_{k=2}^{m} \log k! \left[nP_k(m,d) + \binom{n}{2} P_k(m,d,d) \right] - \log C \ ,$$

where

$$C = \frac{(d!\,)^n}{(nd-1)!\,!} \ .$$

The approximate entropy of the distribution of multigraphs under RSM is given by

$$h(\mathbf{P}) \approx \frac{1}{2} \log \left[(2\pi e)^{r-n} \ \det(\mathbf{\Sigma}_{\rm RSM}) \ \right] \ ,$$

where $\mathbf{\Sigma}_{\rm RSM}$ is the covariance matrix of $r-n$ non-redundant components of the multiplicity sequence $\mathbf{m}$. In order to find $\mathbf{\Sigma}_{\rm RSM}$, consider $Q_{ijk\ell}$ for all $i \le j$ and $k \le \ell$ where $Q_{ijk\ell} = Q_{k\ell ij}$. More specifically, we need the formulae for two loops that are at the same vertex or at different vertices, one loop and one non-loop with one or no common vertex, and two non-loops with two, one or no vertices in common. With a slight abuse of notation, the r by r covariance matrix under RSM can be written as

$$\mathbf{\Sigma}_{\rm RSM} = \mathbf{\Sigma}_{\rm IEA} + \mathbf{\Delta} \ ,$$

where $\mathbf{\Sigma}_{\rm IEA}$ is the r by r covariance matrix under independent edge assignments, i.e. the covariance matrix of a multinomial distribution with parameters m and $\mathbf{Q}$. The elements of $\mathbf{\Sigma}_{\rm IEA}$ are $mQ_{ij}(\delta_{ijk\ell} - Q_{k\ell})$ for r different (i,j) and (k,ℓ) in R, where $\delta_{ijk\ell}$ is equal to 1 if $(i,j) = (k,\ell)$ and 0 otherwise. The matrix $\mathbf{\Delta}$ consists of elements $\Delta_{ijk\ell} = m(m-1)(Q_{ijk\ell} - Q_{ij}Q_{k\ell})$ for r different (i,j) and (k,ℓ) in site space R. Renaming the ordered edge sequence indexes $(1,1) < (1,2) < \cdots < (1,n) < (2,2) < (2,3) < \cdots < (n,n)$ to $1, 2, \ldots, r = \binom{n+1}{2}$, $\mathbf{\Sigma}_{\rm RSM}$ can be written as

$$\mathbf{\Sigma}_{\rm RSM} = m \begin{pmatrix} Q_1(1-Q_1) & -Q_1Q_2 & \cdots & -Q_1Q_r \\ -Q_1Q_2 & Q_2(1-Q_2) & \cdots & -Q_2Q_r \\ \vdots & \vdots & \ddots & \vdots \\ -Q_1Q_r & -Q_2Q_r & \cdots & Q_r(1-Q_r) \end{pmatrix} + m(m-1) \begin{pmatrix} \Delta_{11} & \Delta_{12} & \cdots & \Delta_{1r} \\ \Delta_{12} & \Delta_{22} & \cdots & \Delta_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta_{1r} & \Delta_{2r} & \cdots & \Delta_{rr} \end{pmatrix} .$$

In order to avoid singularity of $\mathbf{\Sigma}_{\text{RSM}}$, remove n of the r components in $\mathbf{m}$ that are linear combinations of the others, implying that the degrees of freedom here is equal to $r - n = \binom{n}{2}$. A similar argument was used in Section 5 for the trivariate distribution of the edge multiplicities (m_{ii}, m_{jj}, m_{ij}) obtained when three of the six pairs of vertex categories were redundant.

We now turn to the distribution of multigraphs under IEA which corresponds to that edges in $\boldsymbol{\xi}$ are independently assigned to sites according to the probability distribution $(P_{ij} : (i,j) \in V^2)$, and edges in $\boldsymbol{\eta}$ are independently assigned to sites according to the probability distribution $\mathbf{Q} = (Q_{ij} : (i,j) \in R)$, where P_{ij} and Q_{ij} are defined as earlier. Thus,

$$P(\boldsymbol{\eta} = \mathbf{y}) = \prod_{k=1}^{m} Q(y_{2k-1}, y_{2k}) = \prod_{i \le j} Q_{ij}^{m_{ij}(y)} \ ,$$

where $Q(i,j) = Q_{ij}$, and $\mathbf{m}(\boldsymbol{\eta})$ is multinomially distributed with parameters m and $\mathbf{Q}$ so that

$$P(\mathbf{m}(\boldsymbol{\eta}) = \mathbf{m}) = \binom{m}{\mathbf{m}} \mathbf{Q}^{\mathbf{m}} = \frac{m!}{\prod_{i \le j} m_{ij}!} \prod_{i \le j} Q_{ij}^{m_{ij}} \ ,$$

which is shortly denoted $B_{\mathbf{m}}(\mathbf{d})$. Note that this implies that $m_{ij}(\boldsymbol{\eta})$ is binomially distributed with parameters m and Q_{ij}, as considered in previous sections. Letting $\mathbf{B} = (B_{\mathbf{m}}(\mathbf{d}) :$ all different $\mathbf{m})$ denote the probability distribution of multigraphs under IEA, the entropy of this distribution is equal to

$$h(\mathbf{B}) = \sum_{\mathbf{m}: B_{\mathbf{m}} > 0} -B_{\mathbf{m}} \log B_{\mathbf{m}} \ .$$

The number of multigraphs $\mathbf{m}$ is not restricted by $\mathbf{d}$ in the same way as for the RSM distribution. Here it is given by the total number of ordered partitions of m into $r = \binom{n+1}{2}$ available sites of vertex pairs for the edges:

$$\#\mathbf{m} = \binom{m + r - 1}{m} \ .$$

This gives an upper bound to the entropy

$$h(\mathbf{B}) \le \log \binom{m + r - 1}{m} \ .$$

The approximate entropy under IEA is given by

$$h(\mathbf{B}) \approx \frac{1}{2} \log \left[(2\pi e)^{r-1} \ \det(\mathbf{\Sigma}_{\text{IEA}}) \ \right] \ ,$$

where $\mathbf{\Sigma}_{\rm IEA}$ now denotes the $(r-1)$ by $(r-1)$ covariance matrix of $\mathbf{m}$ when one of the r components is omitted (to avoid singularity). The determinant of this matrix can be proved to be equal to

$$m^{r-1} \prod_{i \leq j} Q_{ij} \ ,$$

where the product is over all r pairs $(i,j) \in R$. Thus, the approximate entropy under IEA is obtained by

$$h(\mathbf{B}) \approx \frac{1}{2} \log \left[(2\pi e m)^{r-1} \prod_{i \leq j} Q_{ij} \right] \ .$$

In Table 8 we consider the entropies of the distributions of multigraphs under RSM and IEA and the entropy approximations for these distributions with $n = 6$ vertices, $m = 9$ edges and different degree sequences with minimum degree 2. Also shown in Table 8 is the divergence between these two distributions. The total number of graphs under IEA for this example is 10,015,005 and we do not calculate the exact entropies which here are close to their approximations. We see that the approximate entropies under RSM and under IEA are both close to their upper bounds. The exact entropies under RSM are close to the approximate entropies implying that the distributions of multigraphs are fairly well approximated by normal distributions. However, the divergence values indicate a large deviation between the distributions under RSM and IEA. These findings indicate rather flat distributions over very different ranges for RSM and IEA. We note that there are cases in Table 8 when the approximate entropies are larger than the upper bounds to the exact entropies. This occurs for the last three rows of Table 8. It would be a tedious task to investigate this in general by using the formula given in Section 5 for the total number of multigraphs under RSM. We therefore restrict ourselves to a general investigation of IEA by considering the following inequality which holds when the IEA entropy approximation is at most equal to the upper bound to the exact entropy:

$$(2\pi e m)^{\frac{r-1}{2}} \left[\prod_{i \leq j} Q_{ij} \right]^{\frac{1}{2}} \leq \binom{m+r-1}{m} \ .$$

The right hand side can be written as

$$\binom{m+r-1}{m} = \prod_{k=1}^{r-1} \left(1 + \frac{m}{k} \right) \ ,$$

and the inequality can be expressed as giving an upper bound to the geometric mean of the edge probabilities according to

$$\tilde{Q} \leq \left(\frac{G}{2\pi e} \right)^{\frac{r-1}{r}} \ ,$$

Table 8: Entropies of the distributions of multigraphs under random stub matching (RSM), independent edge assignments (IEA) and the entropy approximations for these distributions with $n = 6$ vertices, $m = 9$ edges and different degree sequences with minimum degree 2. Also given is the divergence between these two distributions.

Degree sequence	Number of graphs	Entropy RSM Upper bound	Entropy RSM Exact	Entropy RSM Approx	Entropy IEA Upper bound	Entropy IEA Approx	Divergence
$\mathbf{d} = (8,\ 2,\ 2,\ 2,\ 2,\ 2)$	773	9.59	8.63	7.61	23.26	18.84	8.72
$\mathbf{d} = (7,\ 3,\ 2,\ 2,\ 2,\ 2)$	1210	10.24	9.47	8.98	23.26	20.41	8.90
$\mathbf{d} = (6,\ 4,\ 2,\ 2,\ 2,\ 2)$	1651	10.69	9.97	9.62	23.26	21.15	8.99
$\mathbf{d} = (5,\ 5,\ 2,\ 2,\ 2,\ 2)$	1804	10.82	10.14	9.82	23.26	21.37	9.02
$\mathbf{d} = (6,\ 3,\ 3,\ 2,\ 2,\ 2)$	1914	10.90	10.23	10.25	23.26	21.86	9.07
$\mathbf{d} = (5,\ 4,\ 3,\ 2,\ 2,\ 2)$	2424	11.24	10.60	10.77	23.26	22.45	9.14
$\mathbf{d} = (4,\ 4,\ 4,\ 2,\ 2,\ 2)$	2814	11.46	10.81	11.08	23.26	22.82	9.18
$\mathbf{d} = (5,\ 3,\ 3,\ 3,\ 2,\ 2)$	2857	11.48	10.87	11.40	23.26	23.17	9.21
$\mathbf{d} = (4,\ 4,\ 3,\ 3,\ 2,\ 2)$	3316	11.70	11.08	11.72	23.26	23.53	9.26
$\mathbf{d} = (4,\ 3,\ 3,\ 3,\ 3,\ 2)$	3943	11.95	11.36	12.35	23.26	23.27	9.33
$\mathbf{d} = (3,\ 3,\ 3,\ 3,\ 3,\ 3)$	4720	12.21	11.64	12.98	23.26	24.97	9.41

where

$$\tilde{Q} = \left[\prod_{i\le j} Q_{ij}\right]^{\frac{1}{r}} \quad \text{and} \quad G = \left[\prod_{k=1}^{r-1}\left(\frac{1}{\sqrt{m}} + \frac{\sqrt{m}}{k}\right)^2\right]^{\frac{1}{r-1}} .$$

Under IEA, we have that

$$\begin{aligned}\prod_{i\le j} Q_{ij} &= \frac{d_1(d_1-1)\ d_2(d_2-1)\cdots d_n(d_n-1)\ 2d_1d_2\ 2d_1d_3\cdots 2d_{n-1}d_n}{[2m(2m-1)]^r} \\ &= \frac{2^{\binom{n}{2}}\ (d_1-1)\ (d_2-1)\ \cdots (d_n-1)\ (d_1\ d_2\cdots d_n)^n}{[2m(2m-1)]^r} ,\end{aligned}$$

and the geometric mean is

$$\tilde{Q} = \left[\prod_{i\le j} Q_{ij}\right]^{\frac{1}{r}} = \left[\frac{2^{\binom{n}{2}}\prod_{i=1}^n (d_i-1)\,(\prod_{i=1}^n d_i)^n}{[2m(2m-1)]^r}\right]^{\frac{1}{r}} .$$

A comparison between the distributions $\mathbf{P}$ and $\mathbf{B}$ for regular graphs with $n = 4$ vertices of the same degree d is shown in Figure 12 when d varies from 2 to 10 so that the number of edges $m = nd/2 = 2d$ varies from 4 to 20. There is one case where the approximate entropy

under IEA is larger than the upper bound to the exact entropy under IEA. This occurs for $m = 6$ and $\mathbf{d} = (3\ 3\ 3\ 3)$. We investigate this by using the above shown inequality. For this case we have that $r = 10$ and the geometric mean of the edge probabilities is

$$\tilde{Q} = \left[\frac{2^{\binom{n}{2}}(d-1)^n\,(d)^{n^2}}{[2m(2m-1)]^r}\right]^{\frac{1}{r}} = \left[\frac{2^6 2^4 3^{16}}{[12(11)]^{10}}\right]^{\frac{1}{10}} = 0.088\ .$$

Further, we have that

$$G = \left[\prod_{k=1}^{r-1}\left(\frac{1}{\sqrt{m}} + \frac{\sqrt{m}}{k}\right)^2\right]^{\frac{1}{r-1}} = \left[\prod_{k=1}^{9}\left(\frac{1}{\sqrt{6}} + \frac{\sqrt{6}}{k}\right)^2\right]^{\frac{1}{9}} = 1.107\ ,$$

so that the right hand side of the inequality is equal to

$$\left(\frac{G}{2\pi e}\right)^{\frac{r-1}{r}} = \left(\frac{1.107}{2\pi e}\right)^{\frac{9}{10}} = 0.085\ .$$

As seen, the geometric mean of the edge probabilities is greater than the upper bound which implies that the inequality is not satisifed, i.e. the IEA entropy approximation for this example is greater than the upper bound to the exact entropy.

Further in Figure 12, we note that as the number of edges increases, the differences between the upper bounds of the entropies and the exact or approximate entropies are increased. This indicates that the distributions of multigraphs under RSM and IEA cluster at the high probability sites when more edges are added and therefore are less flat for large values of m.

7 Simplicity and Complexity

The probability distribution of complexity of multigraphs generated by RSM depends in a complicated way on its degree sequence. Different aspects of complexity can be studied by various indicators and summary measures. For instance, the expected value of a simplicity indicator is the probability that the multigraph is simple, and it has received much attention in the literature. Janson (2009), McKay (1985), McKay and Wormald (1991), Bollobàs (1980), and Bender and Canfield (1978) all focus on asymptotic results and so far no exact solution seems to have been found. Other examples of useful information about complexity are given by the expected numbers of loops and multiple edges and their variances. This section reviews some results from the literature and presents some convenient summary measures of complexity.

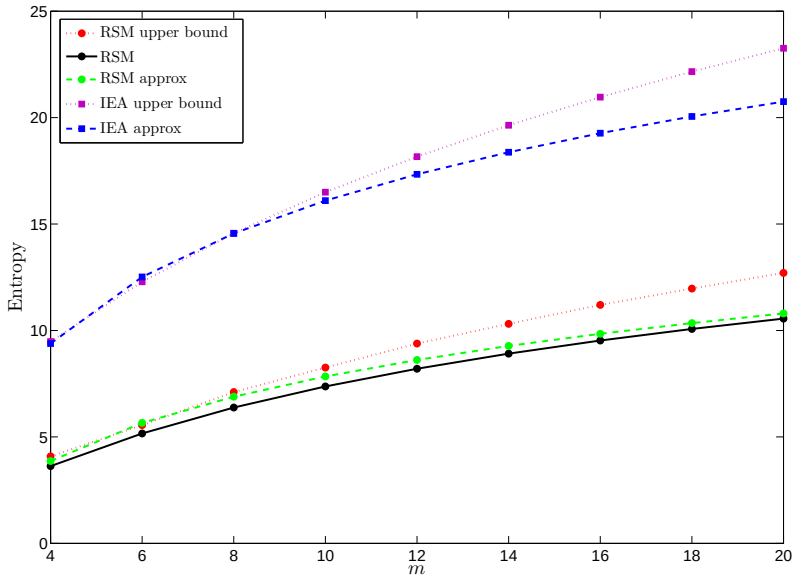


Figure 12: Approximate and exact entropies of the distribution of multigraphs under random stub matching (RSM), approximate entropies under independent edge assignments (IEA) and upper bounds of entropies in different regular multigraphs with $n = 4$ vertices. Entropies are plotted against number of edges m between 4 and 20.

There is a considerable literature about graphical degree sequences, i.e. such degree sequences that can be realized by simple graphs. Obvious necessary conditions for a finite sequence of non-negative integers $(d_1, \ldots, d_n)$ to be graphical is that $d_i < n$, $\sum_{i=1}^{n} d_i = 2m$ is even and $m \leq \binom{n}{2}$. Erdös and Gallai (1960) give further necessary and sufficient conditions for the existence of a simple graph with given degree sequence. They show that a degree sequence $\mathbf{d}$ of non-negative integers in non-increasing order $d_1 \geq \cdots \geq d_n$ is graphical if and only if

$$\sum_{i=1}^{j} d_i \leq j(j-1) + \sum_{i=j+1}^{n} \min(j, d_i), \qquad \text{for } 1 \leq j \leq n-1 \ .$$

Proof of necessity is straightforward; the left side of the inequality counts degree among the j vertices with highest degrees. The first term on the right side is a consequence of that at most $(j-1)$ edges are incident to any of the first j vertices. The second term of the right side is a sum of upper bounds to the number of edges for each remaining vertex. The proof of sufficiency is more complicated but can be found in several papers including Tripathi,

Venugopalan and West (2009), Sierksma and Hoogeveen (1991) and Choudum (1986).

Besides the existence result above, a recursive test to find graphical degree sequences is given by Havel (1955) and Hakimi (1962). To avoid isolated vertices only sequences with strictly positive degrees are considered in their result. They show that a non-increasing sequence $d_1 \geq \cdots \geq d_n \geq 1$ with $n \geq 2$ is graphical if and only if the sequence

$$\mathbf{d}^* = (d_2 - 1, d_3 - 1, \ldots, d_{d_1+1} - 1, d_{d_1+2}, \ldots, d_n)$$

is graphical. The proof, which can be adapted into an algorithm to determine whether or not a sequence of positive integers can be realized by a simple graph, can be found in several papers including Blitzstein and Diaconis (2011) and Tripathi and Tyagi (2008).

The asymptotic results given by Janson (2009) concern the probability that an RSM multigraph is simple, which we denote P_0. The asymptotic probabilities given by Janson (2009) are based on the assumptions that degrees and numbers of vertices and edges depend on some parameter that tends to infinity. The main result is that as $m \to \infty$:

$$(i)\ \liminf P_0 > 0 \text{ if and only if } \sum d_i^2 = O(m)\ ,$$

$$(ii)\ P_0 \to 0 \text{ if and only if } \frac{\sum d_i^2}{m} \to \infty\ ,$$

with the corollary that as $n \to \infty$ where $m = O(n)$ and $n = O(m)$:

$$(i)\ \liminf P_0 > 0 \text{ if and only if } \sum d_i^2 = O(n)\ ,$$

$$(ii)\ P_0 \to 0 \text{ if and only if } \frac{\sum d_i^2}{n} \to \infty\ .$$

The two asymptotic formulas for the probability that a multigraph is simple are given by Janson (2009) and they can in our notations be given as

$$P_0^{'} = \exp\left[-\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_{ij} + \sum_{1\leq i<j\leq n}\sum\log(1+\lambda_{ij})\right] + o(1)\ ,$$

and, assuming that $\max_i(d_i) = o(\sqrt{m})$,

$$P_0^{''} = \exp\left[-\Lambda(1+\Lambda)\right] + o(1)\ ,$$

where

$$\lambda_{ij} = \frac{1}{m}\sqrt{\binom{d_i}{2}\binom{d_j}{2}}$$

and

$$\Lambda = \frac{1}{2}\sum_{i=1}^{n}\lambda_{ii} = \frac{1}{2m}\sum_{i=1}^{n}\binom{d_i}{2} = \frac{1}{4m}\sum_{i=1}^{n}d_i^2 - \frac{1}{2}\ .$$

Hence

$$P_0^{'} = \exp\left[-\frac{1}{2m}\left(\sum_{i=1}^{n}\sqrt{\binom{d_i}{2}}\right)^2 + \sum_{1\le i<j\le n}\sum\log\left(1+\frac{1}{m}\sqrt{\binom{d_i}{2}\binom{d_j}{2}}\right)\right] + o(1)\ ,$$

and

$$P_0^{''} = \exp\left[-\frac{1}{4}\left(\frac{1}{2m}\sum_{i=1}^{n}d_i^2\right)^2 + \frac{1}{4}\right] + o(1)\ .$$

In particular, for regular graphs with the same degree d at every vertex, $\lambda_{ij} = (d-1)/n$ and $\Lambda = (d-1)/2$ so that

$$P_0^{'} = \exp\left[-\frac{n(d-1)}{2} + \binom{n}{2}\log\left(1+\frac{(d-1)}{n}\right)\right] + o(1)$$

when $nd \to \infty$ and $n \to \infty$, and

$$P_0^{''} = \exp\left[-\frac{(d-1)(d+1)}{4}\right] + o(1)$$

when $d/n \to 0$ and $n \to \infty$. Some numerical examples for these approximations are presented later in this section.

Using the results obtained in Section 4 for edge multiplicities under RSM, we derive expected values and variances of some quantities that can be used to study simplicity and complexity of multigraphs. The expected values of the numbers of loops $m_1 = m_1(\boldsymbol{\eta})$ (already mentioned in previous section) and non-loops $m_2 = m_2(\boldsymbol{\eta})$ under RSM are directly obtained as expected values of local multiplicities according to:

$$E(m_1) = m\sum_{i=1}^{n}Q_{ii} = \frac{1}{2m-1}\sum_{i=1}^{n}\binom{d_i}{2}$$

and

$$E(m_2) = m\sum_{i<j}\sum Q_{ij} = \frac{1}{2m-1}\sum_{i<j}\sum d_i d_j\ .$$

We also obtain $E(m_2) = m - E(m_1)$ by using the linear relationship $m_2 = m - m_1$. This linear relationship also implies that

$$\mathrm{Var}(m_2) = \mathrm{Var}(m_1)\ .$$

This common variance is given by

$$\mathrm{Var}(m_1) = \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(m_{ii}, m_{jj}) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{\ell=1}^{m}\mathrm{Cov}(I_{iik}, I_{jj\ell}) \ ,$$

where $\mathrm{Cov}(I_{iik}, I_{jj\ell})$ need to be considered for different i, j, k and ℓ. For $k = \ell$

$$\mathrm{Cov}(I_{iik}, I_{jjk}) = \begin{cases} Q_{ii}(1 - Q_{ii}) & \text{for } i = j \\ -Q_{ii}Q_{jj} & \text{for } i \neq j \ , \end{cases}$$

and for $k \neq \ell$

$$\mathrm{Cov}(I_{iik}, I_{jj\ell}) = \begin{cases} Q_{iiii} - Q_{ii}^2 & \text{for } i = j \\ Q_{iijj} - Q_{ii}Q_{jj} & \text{for } i \neq j \ , \end{cases}$$

where

$$Q_{iiii} = \frac{\binom{d_i}{2}\binom{d_i-2}{2}}{\binom{2m}{2}\binom{2m-2}{2}} \quad \text{and} \quad Q_{iijj} = \frac{\binom{d_i}{2}\binom{d_j}{2}}{\binom{2m}{2}\binom{2m-2}{2}} \ .$$

This implies that

$$\begin{aligned}
\mathrm{Var}(m_1) &= m\left[\sum_{i=1}^{n} Q_{ii}(1 - Q_{ii}) - \sum\sum_{i\neq j} Q_{ii}Q_{jj}\right] \\
&\qquad + m(m-1)\left[\sum_{i=1}^{n}(Q_{iiii} - Q_{ii}^2) + \sum\sum_{i\neq j}(Q_{iijj} - Q_{ii}Q_{jj})\right] \\
&= m\sum_{i=1}^{n} Q_{ii} + m(m-1)\sum_{i=1}^{n} Q_{iiii} - m^2\sum_{i=1}^{n} Q_{ii}^2 \\
&\qquad + m(m-1)\sum\sum_{i\neq j} Q_{iijj} - m^2\sum\sum_{i\neq j} Q_{ii}Q_{jj} \\
&= m\sum_{i=1}^{n} Q_{ii}(1 - m\sum_{i=1}^{n} Q_{ii}) + m(m-1)\left(\sum_{i=1}^{n} Q_{iiii} + \sum\sum_{i\neq j} Q_{iijj}\right) \\
&= \frac{1}{2m-1}\sum_{i=1}^{n}\binom{d_i}{2}\left[1 - \frac{1}{2m-1}\sum_{j=1}^{n}\binom{d_j}{2}\right] \\
&\qquad + \frac{1}{(2m-1)(2m-3)}\left[\sum_{i=1}^{n}\binom{d_i}{2}\binom{d_i-2}{2} + \sum\sum_{i\neq j}\binom{d_i}{2}\binom{d_j}{2}\right] \ .
\end{aligned}$$

In particular, for regular graphs with the same degree d at every vertex we obtain

$$E(m_1) = \frac{d-1}{2}\left(1 + \frac{1}{nd-1}\right)$$

and

$$\mathrm{Var}(m_1) = \frac{d-1}{2}\left(1 + \frac{1}{nd-1} + \frac{(d-2)(d-3)}{2nd}\right) + O\left(\frac{1}{n^2}\right) .$$

Hence we expect that there are slightly more than $(d-1)/2$ loops, and the expected number of loops is about the same for any number of vertices. The variance indicates that the number of loops might be approximately Poisson distributed.

A variable that has been used by several authors to study simplicity is $m_1 + m_3$ where m_3 is the number of pairs of equal non-loops in $\boldsymbol{\eta}$. This number is formally given by

$$m_3 = \sum_{i<j}\sum \binom{m_{ij}}{2} = \sum_{i<j}\sum_{k<\ell}\sum\sum I_{ijk} I_{ij\ell} .$$

The sum $m_1 + m_3$ is a variable that is 0 if and only if the multigraph is simple. Now

$$E(m_3) = \frac{m(m-1)}{2}\sum_{i<j}\sum Q_{ijij} = \frac{2}{(2m-1)(2m-3)}\sum_{i<j}\sum \binom{d_i}{2}\binom{d_j}{2} ,$$

and the expected value of $m_1 + m_3$ is thus given by

$$\begin{aligned} E(m_1 + m_3) &= m\sum_{i=1}^{n} Q_{ii} + \binom{m}{2}\sum_{i<j}\sum Q_{ijij} \\ &= \frac{1}{2m-1}\sum_{i=1}^{n}\binom{d_i}{2} + \frac{2}{(2m-1)(2m-3)}\sum_{i<j}\sum \binom{d_i}{2}\binom{d_j}{2} . \end{aligned}$$

In particular, for regular graphs with the same degree d at every vertex this expected value is about $E(m_1 + m_3) = (d^2-1)/4$ regardless of the number of vertices.

In order to make it easier to investigate simplicity and complexity, and find the expected value of the number $r_k = r_k(\boldsymbol{\eta})$ of sites with occupancy k, we use the IEA multiplicity distribution introduced in the previous section so that $\mathbf{m}(\boldsymbol{\eta})$ is multinomially distributed with parameters m and $\mathbf{Q}$. From this it follows that $m_{ij}(\boldsymbol{\eta})$ is binomially distributed with parameters m and Q_{ij} and

$$E(r_k) = \sum_{i\le j}\sum \binom{m}{k} Q_{ij}^k (1-Q_{ij})^{m-k} \quad \text{for } k = 0, 1, \ldots, m .$$

Using this result, we obtain a formula for the expected value of the statistic $t = t(\boldsymbol{\eta})$, which in Section 3 was shown to determine the probability distribution of multigraphs under RSM. This statistic is a summary measure of complexity that is equal to 0 if and only if the multigraph is simple. Its expected value under IEA is given by

$$\begin{aligned} E(t) &= E\left[m_1 + \sum_{k=2}^{m} r_k \log k!\right] \\ &= m\sum_{i=1}^{n} Q_{ii} + \sum_{k=2}^{m}\left[\log k! \binom{m}{k} \sum\sum_{i\le j} Q_{ij}^k (1-Q_{ij})^{m-k}\right] \ . \end{aligned}$$

Note that the expected value of t under RSM that was given in the previous section is considerably more complicated to specify and analyze.

Other statistics related to complexity are also easily handled under IEA. The number of sites with no occupancy given by $r_0 = r_0(\boldsymbol{\eta})$ has expected value

$$E(r_0) = \sum\sum_{i\le j} (1-Q_{ij})^m \ ,$$

and the number of sites with single occupancy given by $r_1 = r_1(\boldsymbol{\eta})$ has expected value

$$E(r_1) = m\sum\sum_{i\le j} Q_{ij}(1-Q_{ij})^{m-1} \ .$$

The expected value of the number of multiple occupancy sites is thus

$$E(r - r_0 - r_1) = r - \sum\sum_{i\le j} (1-Q_{ij})^m - m\sum\sum_{i\le j} Q_{ij}(1-Q_{ij})^{m-1} \ ,$$

and the number of multiple edges has expected value

$$E(m - r_1) = m\left(1 - \sum\sum_{i\le j} Q_{ij}(1-Q_{ij})^{m-1}\right) \ .$$

A particulary interesting statistic is r_{21} (defined in Section 3) which is equal to m if and only if the multigraph is simple. This means that there are m single occupancies of non-loops. The exact probability distribution of this statistic is unknown, but being a counting statistic it is not unreasonable to assume that it is approximately Poisson distributed with parameter $\lambda = E(r_{21})$. The probability of simplicity $P_0 = P(r_{21} = m)$ can then be approximated by

$$P_0''' = \frac{e^{-\lambda}\lambda^m}{m!} \ ,$$

where

$$\lambda = E(r_{21}) = m \sum_{i \leq j} \sum Q_{ij}(1 - Q_{ij})^{m-1} = m \sum_{i<j} \sum \frac{d_i d_j}{m(2m-1)} \left(1 - \frac{d_i d_j}{m(2m-1)}\right)^{m-1} .$$

For regular graphs with the same degree d at every vertex we obtain

$$E(r_{21}) = \frac{\binom{n}{2} d^2}{(nd-1)} \left(1 - \frac{2d}{n(nd-1)}\right)^{\frac{nd}{2}-1} .$$

This expected value is for large n and small d approximately equal to

$$E(r_{21}) \approx \frac{\binom{n}{2} d^2}{(nd-1)} \exp\left(-\frac{d(nd-2)}{n(nd-1)}\right) \approx \frac{(n-1)d}{2} \exp\left(-\frac{d}{n}\right) \approx \frac{nd}{2} - \frac{d(d+1)}{2} + \frac{d^2(d+2)}{4n} ,$$

and it follows for this case that

$$P_0''' = \frac{e^{\lambda} \lambda^m}{m!} \approx \frac{\lambda^m e^{m-\lambda}}{m^m \sqrt{2\pi m}} \approx \frac{e^{\binom{d+1}{2}}}{\sqrt{2\pi m}} \left(1 - \frac{\binom{d+1}{2}}{m}\right)^m .$$

In Table 9 we give some numerical examples of the probability that a RSM multigraph is simple. These probabilities are compared to the previously given asymptotic probabilities P_0' and P_0'' and the approximate probability P_0'''. Here, we look at small graphs with 6 to 8 vertices, and study cases where the numbers of edges are in the interval ± 1 of the number of vertices. For each case presented in Table 9, we focus on degree sequences with positive degrees at each vertex that are at most 3 so that a reasonable number of simple graphs is possible. From Table 9 we see that the Poisson approximation is close to the RSM probability of simplicity but, as expected, the asymptotic probabilites do not perform well for these small examples. In particular, the best Poisson approximations are for cases where $m = n + 1$. Further we note that the Poisson approximations do not perform well for the regular graphs presented in Table 9.

Table 9: Some numerical examples of the probability that an RSM multigraph is simple, compared to the suggested Poisson approximation of m single edge occupancies of non-loops, and the two asymptotic probabilities suggested by Janson (2009).

		$n = 6$, $m = 5$		
Degree sequence	RSM	Poisson	Asymptotic 1	Asymptotic 2
$\mathbf{d} = (3,\ 3,\ 1,\ 1,\ 1,\ 1)$	0.107	0.104	0.482	0.383
$\mathbf{d} = (3,\ 2,\ 2,\ 1,\ 1,\ 1)$	0.195	0.116	0.540	0.472
$\mathbf{d} = (2,\ 2,\ 2,\ 2,\ 1,\ 1)$	0.284	0.128	0.603	0.571
		$n = 6$, $m = 6$		
Degree sequence	RSM	Poisson	Asymptotic 1	Asymptotic 2
$\mathbf{d} = (3,\ 3,\ 3,\ 1,\ 1,\ 1)$	0.042	0.070	0.356	0.269
$\mathbf{d} = (3,\ 3,\ 2,\ 2,\ 1,\ 1)$	0.086	0.082	0.401	0.329
$\mathbf{d} = (3,\ 2,\ 2,\ 2,\ 2,\ 1)$	0.132	0.093	0.450	0.397
$\mathbf{d} = (2,\ 2,\ 2,\ 2,\ 2,\ 2)$	0.180	0.104	0.503	0.472
		$n = 6$, $m = 7$		
Degree sequence	RSM	Poisson	Asymptotic 1	Asymptotic 2
$\mathbf{d} = (3,\ 3,\ 3,\ 3,\ 1,\ 1)$	0.031	0.051	0.276	0.204
$\mathbf{d} = (3,\ 3,\ 3,\ 2,\ 2,\ 1)$	0.047	0.061	0.311	0.246
$\mathbf{d} = (3,\ 3,\ 2,\ 2,\ 2,\ 2)$	0.069	0.070	0.349	0.294
		$n = 7$, $m = 6$		
Degree sequence	RSM	Poisson	Asymptotic 1	Asymptotic 2
$\mathbf{d} = (3,\ 3,\ 2,\ 1,\ 1,\ 1,\ 1)$	0.132	0.100	0.473	0.397
$\mathbf{d} = (3,\ 2,\ 2,\ 2,\ 1,\ 1,\ 1)$	0.200	0.110	0.526	0.472
$\mathbf{d} = (2,\ 2,\ 2,\ 2,\ 2,\ 1,\ 1)$	0.270	0.119	0.582	0.554
		$n = 7$, $m = 7$		
Degree sequence	RSM	Poisson	Asymptotic 1	Asymptotic 2
$\mathbf{d} = (3,\ 3,\ 3,\ 2,\ 1,\ 1,\ 1)$	0.068	0.076	0.365	0.294
$\mathbf{d} = (3,\ 3,\ 2,\ 2,\ 2,\ 1,\ 1)$	0.103	0.085	0.406	0.348
$\mathbf{d} = (3,\ 2,\ 2,\ 2,\ 2,\ 2,\ 1)$	0.143	0.093	0.451	0.407
$\mathbf{d} = (2,\ 2,\ 2,\ 2,\ 2,\ 2,\ 2)$	0.189	0.102	0.499	0.472
		$n = 7$, $m = 8$		
Degree sequence	RSM	Poisson	Asymptotic 1	Asymptotic 2
$\mathbf{d} = (3,\ 3,\ 3,\ 3,\ 2,\ 1,\ 1)$	0.043	0.059	0.291	0.229
$\mathbf{d} = (3,\ 3,\ 3,\ 2,\ 2,\ 2,\ 1)$	0.060	0.067	0.324	0.269
$\mathbf{d} = (3,\ 3,\ 2,\ 2,\ 2,\ 2,\ 2)$	0.082	0.075	0.360	0.313
		$n = 8$, $m = 7$		
Degree sequence	RSM	Poisson	Asymptotic 1	Asymptotic 2
$\mathbf{d} = (3,\ 3,\ 3,\ 1,\ 1,\ 1,\ 1,\ 1)$	0.098	0.090	0.424	0.348
$\mathbf{d} = (3,\ 3,\ 2,\ 2,\ 1,\ 1,\ 1,\ 1)$	0.148	0.098	0.469	0.407
$\mathbf{d} = (3,\ 2,\ 2,\ 2,\ 2,\ 1,\ 1,\ 1)$	0.202	0.106	0.516	0.472
$\mathbf{d} = (2,\ 2,\ 2,\ 2,\ 2,\ 2,\ 1,\ 1)$	0.262	0.113	0.566	0.542
		$n = 8$, $m = 8$		
Degree sequence	RSM	Poisson	Asymptotic 1	Asymptotic 2
$\mathbf{d} = (3,\ 3,\ 3,\ 3,\ 1,\ 1,\ 1,\ 1)$	0.059	0.071	0.337	0.269
$\mathbf{d} = (3,\ 3,\ 3,\ 2,\ 2,\ 1,\ 1,\ 1)$	0.084	0.079	0.373	0.313
$\mathbf{d} = (3,\ 3,\ 2,\ 2,\ 2,\ 2,\ 1,\ 1)$	0.115	0.086	0.411	0.362
$\mathbf{d} = (3,\ 2,\ 2,\ 2,\ 2,\ 2,\ 2,\ 1)$	0.151	0.093	0.452	0.415
$\mathbf{d} = (2,\ 2,\ 2,\ 2,\ 2,\ 2,\ 2,\ 2)$	0.193	0.099	0.496	0.472
		$n = 8$, $m = 9$		
Degree sequence	RSM	Poisson	Asymptotic 1	Asymptotic 2
$\mathbf{d} = (3,\ 3,\ 3,\ 3,\ 3,\ 1,\ 1,\ 1)$	0.040	0.058	0.275	0.217
$\mathbf{d} = (3,\ 3,\ 3,\ 3,\ 2,\ 2,\ 1,\ 1)$	0.053	0.065	0.305	0.251
$\mathbf{d} = (3,\ 3,\ 3,\ 2,\ 2,\ 2,\ 2,\ 1)$	0.071	0.071	0.336	0.288
$\mathbf{d} = (3,\ 3,\ 2,\ 2,\ 2,\ 2,\ 2,\ 2)$	0.093	0.078	0.369	0.329

A convenient way to obtain the IEA multiplicity distribution is to assume that the stubs are randomly generated and independently assigned to vertices, independent stub assignments (ISA). If stubs ξ_k for $k = 1, ..., 2m$ are independently and identically distributed according to a probability distribution $\mathbf{p} = (p_1, ..., p_n)$ with positive probabilities for the n vertices, it follows that the sequence of stub frequencies $\mathbf{d}(\boldsymbol{\xi})$ is multinomially distributed with parameters $2m$ and $\mathbf{p}$. It also follows that edges in $\boldsymbol{\xi}$ are independent and equal to (i, j) with probabilities $P_{ij} = p_i p_j$ for $i = 1, \ldots, n$ and $j = 1, \ldots, n$. Edges in $\boldsymbol{\eta}$ are independent and equal to (i, j) with probabilities $Q_{ij} = p_i^2$ for $i = j$, and $Q_{ij} = 2p_i p_j$ for $i < j$. Now the edge multiplicity sequence $\mathbf{m}(\boldsymbol{\eta})$ is multinomially distributed with parameters m and $\mathbf{Q}$. This is an IEA distribution with a new $\mathbf{Q}$ based on ISA. The conditional distribution of $\mathbf{m}(\boldsymbol{\eta})$ given $\mathbf{d}(\boldsymbol{\xi})$ is equal to the previous edge multiplicity distribution obtained by random stub matching with fixed degrees. This is a consequence of that the conditional probabilities under ISA and IEA can be transformed according to

$$P(\mathbf{m}(\boldsymbol{\eta}) = \mathbf{m} | \mathbf{d}(\boldsymbol{\xi}) = \mathbf{d}) = \frac{\binom{m}{\mathbf{m}} \mathbf{Q}^{\mathbf{m}}}{\binom{2m}{\mathbf{d}} \mathbf{p}^{\mathbf{d}}} = \frac{\binom{m}{\mathbf{m}} 2^{m_2}}{\binom{2m}{\mathbf{d}}} \ ,$$

using that

$$\mathbf{Q}^{\mathbf{m}} = \prod_{i \leq j} Q_{ij}^{m_{ij}} = \left(\prod_{i=1}^{n} p_i^{2m_{ii}} \right) \left(\prod_{i<j} (2p_i p_j)^{m_{ij}} \right) = 2^{m_2} \prod_{i=1}^{n} p_i^{d_i} = 2^{m_2} \mathbf{p}^{\mathbf{d}}$$

and $m_2 = \sum \sum_{i<j} m_{ij}$. The multinomial distribution for $\mathbf{d}(\boldsymbol{\xi})$ with parameters $2m$ and $\mathbf{p}$ can be considered as a Bayesian model for the stub frequencies.

By using that the sequence of stub frequencies $\mathbf{d}(\boldsymbol{\xi})$ is multinomially distributed with parameters $2m$ and $\mathbf{p}$ under ISA, we derive a formula for the expected entropy of the distribution of multigraphs $\mathbf{m}(\boldsymbol{\eta})$. This expected value is found by using the expected entropy under ISA which is equal to the difference $H(\mathbf{m}) - H(\mathbf{d})$ using calculation rules for entropy (given for instance in Frank 2011). Under ISA we use normal approximations to the multinomial distributions of $\mathbf{m}$ and $\mathbf{d}$ and obtain the approximate entropies

$$H(\mathbf{m}) \approx \log \sqrt{(2\pi e m)^{r-1} \prod_{i \leq j} Q_{ij}}$$

and

$$H(\mathbf{d}) \approx \log \sqrt{(4\pi e m)^{n-1} \prod_{i} p_i} \ .$$

It follows that the expected entropy under ISA is given by

$$\begin{aligned} E\left[H(\mathbf{m}|\mathbf{d})\right] &\approx \log \sqrt{(2\pi e m)^{r-1} \prod_{i\le j} Q_{ij}} - \log \sqrt{(4\pi e m)^{n-1} \prod_{i=1}^{n} p_i} \\ &= \log \sqrt{\frac{(2\pi e m)^{r-n} \prod_{i\le j} Q_{ij}}{2^{n-1} \prod_{i=1}^{n} p_i}} \\ &= \log \sqrt{\frac{(2\pi e m)^{\binom{n}{2}} 2^{\binom{n}{2}} (p_1 \cdots p_n)^{n+1}}{2^{n-1} (p_1 \cdots p_n)}} \\ &= \log \sqrt{(2\pi e m)^{\binom{n}{2}} 2^{\binom{n-1}{2}} (p_1 \cdots p_n)^{n}} \ . \end{aligned}$$

For fixed n and $\mathbf{p}$ this is a linear expression in $\log m$ which is denoted

$$H^* = a^* + b^* \log m \ ,$$

where

$$a^* = a^*(n, \mathbf{p}) = \log \sqrt{(2\pi e)^{\binom{n}{2}} 2^{\binom{n-1}{2}} (p_1 \cdots p_n)^{n}}$$

and

$$b^* = b^*(n) = n(n-1)/4 \ .$$

If this expected entropy is calculated with $\mathbf{p} = \mathbf{d}/2m$ for a fixed degree sequence $\mathbf{d}$, it can be considered as an approximation to the entropy H of the distribution of multigraphs $\mathbf{m}$ under RSM with this degree sequence $\mathbf{d}$. In particular, for the distribution of multigraphs under RSM with all $d_i = 2m/n$, its entropy is approximately given by

$$H^* = \log \sqrt{(2\pi e m)^{\binom{n}{2}} 2^{\binom{n-1}{2}} n^{-n^2}} \ ,$$

corresponding to uniform $\mathbf{p}$.

Another approximation H^{**} to the entropy H of the distribution of multigraphs $\mathbf{m}$ under RSM can be based on asymptotic results for the expected entropy, $E\left[H(\mathbf{m}|\mathbf{d})\right] = H(\mathbf{m}) - H(\mathbf{d})$, under ISA. Both $\mathbf{m}$ and $\mathbf{d}$ are multinomially distributed and obtained from sequences of m independent identically distributed sites, and $2m$ independent identically distributed stubs. The central limit theorem in weak form yields asymptotic equipartition properties for r_c and n_c stubs. Here r_c and n_c are given by the entropies of site and stub probabilities according to

$$\log r_c = h(\mathbf{Q}) \quad \text{and} \quad \log n_c = h(\mathbf{p}) \ .$$

In particular, for uniform ISA it follows that $n_c = n$ and $r_c = n^2 2^{-(n-1)/n}$ which is about $\binom{n+1}{2}$ for large n. See for instance Cover and Thomas (1991) which has a detailed presentation of the equipartition property. The equipartition properties implies that $\mathbf{m}$ and $\mathbf{d}$ are

asymptotically multinomially distributed with r_c equal site probabilities and n_c equal stub probabilities. Under ISA it holds that

$$h(\mathbf{Q}) = 2h(\mathbf{p}) - 1 + \sum_{i=1}^{n} p_i^2 \ ,$$

and hence

$$r_c = n_c^2 \ 2^{-(n_c-1)/n_c} \ .$$

The asymptotic entropies of $\mathbf{m}$ and $\mathbf{d}$ under ISA are thus given by

$$H_c(\mathbf{m}) = \log \sqrt{(2\pi e m)^{r_c-1} \ r_c^{-r_c}}$$

and

$$H_c(\mathbf{d}) = \log \sqrt{(4\pi e m)^{n_c-1} \ n_c^{-n_c}} \ .$$

The difference

$$H_c(\mathbf{m}) - H_c(\mathbf{d}) = \log \sqrt{\frac{(2\pi e m)^{r_c-1} \ n_c^{n_c}}{(4\pi e m)^{n_c-1} \ r_c^{r_c}}}$$

is the asymptotic expected conditional entropy of $\mathbf{m}$ given $\mathbf{d}$, when $\mathbf{d}$ is obtained from $\mathbf{p}$. The RSM entropy H of $\mathbf{m}$ with a fixed $\mathbf{d}$ can be approximated by the asymptotic expression obtained with $\mathbf{p} = \mathbf{d}/2m$ for this fixed $\mathbf{d}$. Let H^{**} denote this approximation. It can be given as

$$H^{**} = a^{**} + b^{**} \log m \ ,$$

where

$$a^{**} = a^{**}(n, h(\mathbf{p})) = \log \sqrt{\frac{(2\pi e)^{r_c-1} \ n_c^{n_c}}{(4\pi e)^{n_c-1} \ r_c^{r_c}}} \ ,$$

and

$$b^{**} = b^{**}(n, h(\mathbf{p})) = \frac{r_c - n_c}{2} \ .$$

Now $r_c = n^2 2^{-(n_c-1)/n_c}$ where $n_c = 2^{h(\mathbf{p})}$ and $\mathbf{p}$ is choosen as $\mathbf{p} = \mathbf{d}/2m$ in order to estimate the RSM entropy H with this $\mathbf{d}$. Note that the dependency of H^{**} on $\mathbf{p}$ is only via the degree distribution entropy $h(\mathbf{p})$.

Using these results, a few examples are given to illustrate the performance of the approximations H^* and H^{**} to the entropy H of the distribution of multigraphs under RSM. In Table 10 we compare two approximations using multigraphs with $n = 3$ vertices and $m = 6$ edges. Thus, the approximations for this case are given by

$$H^* = \log \sqrt{(\pi e)^3 \ 2 \ (d_1 \cdot d_2 \cdot d_3)^3 (12)^{-6}}$$

and

$$H^{**} = \log \sqrt{\frac{(2\pi e 6)^{r_c - 1}\ n_c^{n_c}}{(4\pi e 6)^{n_c - 1}\ r_c^{r_c}}}\ ,$$

where $\mathbf{d} = (d_1,\ d_2,\ d_3) = 2m\mathbf{p}$ is varied. As seen in Table 10, both approximations are good for degree distributions that are uniformly or nearly uniformly distributed. Note that the approximation H^* is not useful for very skew degree distributions, e.g. $\mathbf{d} = (10,\ 1,\ 1)$, which yields a negative entropy approximation. However, H^{**} yields better approximations for these skew degree distributions.

Table 10: The entropy of the distribution of multigraphs under random stub matching (RSM) and its approximations in multigraphs with $n = 3$ vertices and $m = 6$ edges for various degree sequences $\mathbf{d}$.

$\mathbf{d} = 2m\mathbf{p}$	Entropy RSM ($\mathbf{d}$) H	Expected entropy ISA ($\mathbf{p}$) H^*	Asymptotic entropy ISA ($\mathbf{p}$) H^{**}
(10, 1, 1)	0.440	-0.631	0.754
(9, 2, 1)	1.096	0.641	1.247
(8, 3, 1)	1.651	1.264	1.665
(7, 4, 1)	2.044	1.597	1.965
(6, 5, 1)	2.242	1.747	2.121
(7, 3, 2)	2.362	2.475	2.343
(6, 4, 2)	2.723	2.763	2.642
(5, 5, 2)	2.847	2.852	2.744
(6, 3, 3)	2.889	3.019	2.815
(5, 4, 3)	3.172	3.247	3.057
(4, 4, 4)	3.330	3.386	3.197

To visualize these findings further, we conclude this section with comparisons between the entropy under RSM and the approximations. First, consider regular multigraphs with $n = 4$ vertices having the same degree d that varies from 1 to 15, so that the number of edges $m = nd/2 = 2d$ varies from 2 to 30. Thus, the degree sequences of these multigraphs are uniformly distributed under ISA with $p_i = 1/n$. In Figure 13 where we see that the entropy under RSM is well approximated by both $H^* = a^* + b^* \log m = -2.22 + 3 \log m$, and $H^{**} = a^{**} + b^{**} \log m = -1.67 + 2.76 \log m$. The RSM entropy deviates slightly from linearity in $\log m$, which is easier to see in Figure 14 where the entropy and its approximations are plotted against $\log m$.

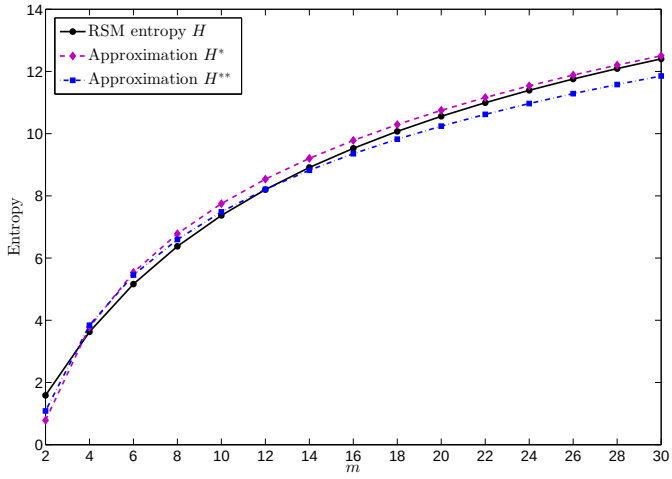


Figure 13: The entropy of the distribution of multigraphs under random stub matching (RSM) and its approximations for different regular multigraphs with $n = 4$ vertices and different numbers of edges m between 2 and 30.

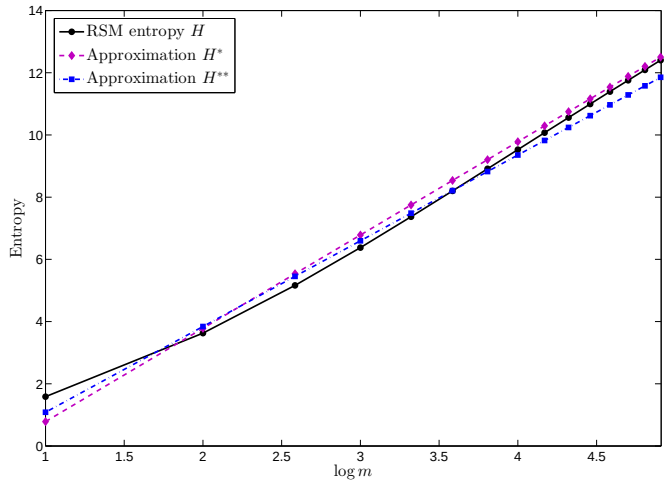


Figure 14: The entropy of the distribution of multigraphs under random stub matching (RSM) and its approximations for different regular multigraphs with $n = 4$ vertices and different numbers of edges m between 2 and 30. Entropy is plotted against $\log m$.

In Figure 15 and 16, we consider two cases with skew degree distributions in multigraphs with $n = 3$ and $n = 4$ vertices. Here, $\mathbf{p} = (4/7,\ 2/7,\ 1/7)$ and $\mathbf{p} = (5/10,\ 3/10,\ 1/10,\ 1/10)$ which implies that possible degree sequences for RSM are multiples of the degree sequences $\mathbf{d} = (8,\ 4,\ 2)$ and $\mathbf{d} = (5,\ 3,\ 1,\ 1)$, respectively. As seen from these two figures, the RSM entropies are well approximated by H^*, but not by H^{**} which deviates more and more from RSM entropy as m increases. When m increases, the same $\mathbf{p}$ will result in different degree sequences $\mathbf{d}$ which in turn give entropies that are harder to approximate by the asymptotic method. From these examples we also note that both the approximations can be either larger or smaller than the RSM entropy H.

In Table 11 we compare the RSM entropy and its approximations using multigraphs with $n = 8$ vertices and $m = 8$ edges. It is clear that approximations H^{**} perform much better than approximations H^*, in particular for the skew degree distributions. However, as the degree sequences become more uniformly distributed, the performance of H^* is improved.

The findings from these considered cases can be summarized as follows. The approximation H^* performs well for multigraphs with degree distributions that are uniformly or close to uniformly distributed. This holds for all multigraphs, no matter size. For skew degree distributions, H^* performs well if the edge frequency m is large. The approximation H^{**} performs well for small multigraphs, i.e. multigraphs with small numbers of vertices and edges, no matter degree distributions. Further, if the number of vertices n increases, this approximation is much better than H^*. If the degree distribution is uniformly distributed, H^{**} also performs well for small number of vertices but large edge frequencies.

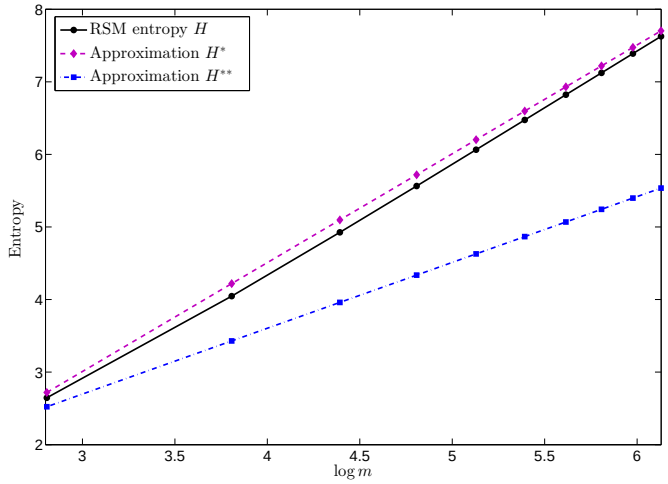


Figure 15: The entropy of the distribution of multigraphs under random stub matching (RSM) and its approximations for multigraphs with $n = 3$ vertices and degree sequences that are multiples of $\mathbf{d} = (8,\ 4,\ 2)$ for edge frequencies $m = 7, 14, 21, \ldots, 70$. Entropy is plotted against $\log m$.

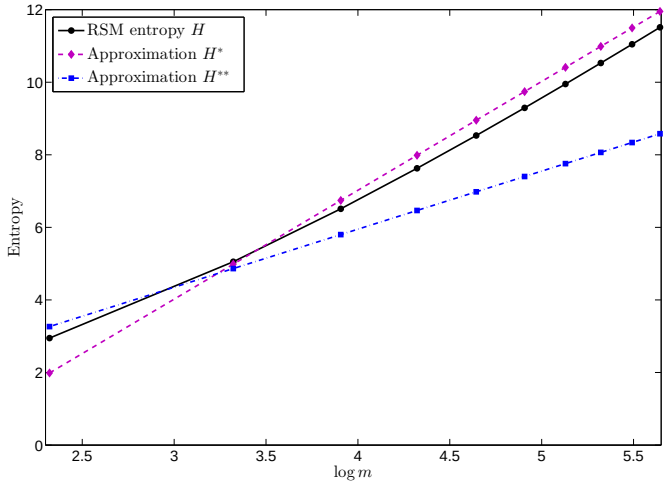


Figure 16: The entropy of the distribution of multigraphs under random stub matching (RSM) and its approximations for multigraphs with $n = 4$ vertices and degree sequences that are multiples of $\mathbf{d} = (5,\ 3,\ 1,\ 1)$ for edge frequencies $m = 5, 10, 15, \ldots, 50$. Entropy is plotted against $\log m$.

Table 11: The entropy of the distribution of multigraphs under random stub matching (RSM) and its approximations in multigraphs with $n = 8$ vertices and $m = 8$ edges for various degree sequences $\mathbf{d}$.

$\mathbf{d} = 2m\mathbf{p}$	Entropy RSM ($\mathbf{d}$) H	Expected entropy ISA ($\mathbf{p}$) H^*	Asymptotic entropy ISA ($\mathbf{p}$) H^{**}
(9, 1, 1, 1, 1, 1, 1, 1)	6.589	-5.502	8.447
(8, 2, 1, 1, 1, 1, 1, 1)	7.869	-2.181	9.897
(7, 3, 1, 1, 1, 1, 1, 1)	8.790	-0.612	10.82
(6, 4, 1, 1, 1, 1, 1, 1)	9.352	0.1589	11.340
(5, 5, 1, 1, 1, 1, 1, 1)	9.541	0.394	11.505
(7, 2, 2, 1, 1, 1, 1, 1)	9.144	1.048	11.259
(6, 3, 2, 1, 1, 1, 1, 1)	9.992	2.498	12.037
(5, 4, 2, 1, 1, 1, 1, 1)	10.419	3.106	12.391
(5, 3, 3, 1, 1, 1, 1, 1)	10.701	3.786	12.646
(4, 4, 3, 1, 1, 1, 1, 1)	10.938	4.159	12.831
(6, 2, 2, 2, 1, 1, 1, 1)	10.379	4.159	12.443
(5, 3, 2, 2, 1, 1, 1, 1)	11.109	5.446	13.023
(4, 4, 2, 2, 1, 1, 1, 1)	11.352	5.819	13.195
(4, 3, 3, 2, 1, 1, 1, 1)	11.652	6.498	13.417
(3, 3, 3, 3, 1, 1, 1, 1)	11.958	7.178	13.625
(5, 2, 2, 2, 2, 1, 1, 1)	11.526	7.106	13.373
(4, 3, 2, 2, 2, 1, 1, 1)	12.084	8.159	13.731
(3, 3, 3, 2, 2, 1, 1, 1)	12.398	8.838	13.914
(4, 2, 2, 2, 2, 2, 1, 1)	12.525	9.819	14.005
(3, 3, 2, 2, 2, 2, 1, 1)	12.847	10.498	14.159
(3, 2, 2, 2, 2, 2, 2, 1)	13.304	12.159	14.351
(2, 2, 2, 2, 2, 2, 2, 2)	13.771	13.819	14.482

References

Bayati, M., Kim, J.H. and Saberi, A. (2010), A Sequential Algorithm for Generating Random Graphs, *Algorithmica*, **58**, 860–910.

Bender, E. A. and Canfield, E. R. (1978), The Asymptotic Number of Labeled Graphs with Given Degree Sequences, *Journal of Combinatorial Theory Series A*, **24(3)**, 296–307.

Blitzstein, J. and Diaconis, P. (2011), A Sequential Importance Sampling Algorithm for Generating Random Graphs with Prescribed Degrees, *Internet Mathematics*, **6(4)**, 489–522.

Bollobàs, B. (1980), A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs, *European Journal of Combinatorics*, **1(4)**, 311–316.

Bollobàs, B. (2001), *Random Graphs*, Second Edition, Cambridge: Cambridge University Press.

Britton, T., Deijfen, M. and Martin-Löf A. (2006), Generating Simple Random Graphs with Prescribed Degree Distribution, *Journal of Statistical Physics*, **124(6)**, 1377-1397.

Choudum, S. A. (1986), A Simple Proof of the Erdös-Gallai Theorem on Graph Sequences, *Bulletin of the Australian Mathematical Society*, **33(1)**, 67–70.

Chung, F. and Lu, L. (2002), Connected Components in Random Graphs with Given Expected Degree Sequences, *Annals of Combinatorics*, **6**, 125–145.

Cover, T. and Thomas, J. (1991), *Elements of Information Theory*, New York: Wiley Series in Communication.

Erdös, P. and Gallai, T. (1960), Graphen mit Punkten Vorgeschriebenen Grades, *Matematikai Lapok*, **11**, 264–274.

Frank, O. (2011), Statistical Information Tools for Multivariate Discrete Data, in *Modern Mathematical Tools and Techniques in Capturing Complexity*, eds. L. Pardo, N. Balakrishnan and M. Ángeles Gil, Berlin: Springer Verlag, 177–190.

Frank, O. and Nowicki, K. (1989), On Entropies of Occupancy Distributions, in *Combinatorics and Graph Theory*, eds. Zdzislaw Skupien, Mieczyslaw Borowiecki, Banach Center Publications, **25**, PWN-Polish Scientific Publishers, Warsaw.

Frank, O. and Shafie, T. (2012), Complexity of Families of Multigraphs, to appear in *JSM Proceedings*, Section on Statistical Graphics, Alexandria, VA: American Statistical Association.

Hakimi, S. L. (1962), On Realizability of a Set of Integers as Degrees of the Vertices of a Linear Graph I, *Journal of the Society for Industrial and Applied Mathematics*, **10(3)**, 496–506.

Havel, V. (1955), A Remark on the Existence of Finite Graphs, *Casopis Pest. Mat.*, **80**, 477–480.

Janson, S. (2009), The Probability that a Random Multigraph is Simple, *Combinatorics, Probability and Computing*, **18(1–2)**, 205–225.

McKay B. D. (1985), Asymptotics for Symmetric 0-1 Matrices with Prescribed Row Sums, *Ars Combinatoria*, **19A**, 15–25.

McKay, B. D. and Wormald, N. C. (1991), Asymptotic Enumeration by Degree Sequence of Graphs with degrees $o(n^{1/2})$, *Combinatorica*, **11(4)**, 369–382.

Sierksma, G. and Hoogeveen, H. (1991), Seven Criteria for Integer Sequences being Graphic, *Journal of Graph Theory*, **15(2)**, 223–231.

Tripathi, A. and Tyagi, H. (2008), A Simple Criterion on Degree Sequences of Graphs, *Discrete Applied Mathematics*, **156(18)**, 3513–3517.

Tripathi, A., Venugopalanb, A. and West, D. B. (2010), A Short constructive proof of the Erdös-Gallai Characterization of Graphic Lists, *Discrete Mathematics*, **310(4)**, 833–834.